Polish verbal aspect and its Finnish statistical correlates in the light of a parallel corpus

POLISH VERBAL ASPECT AND ITS FINNISH STATISTICAL CORRELATES IN THE LIGHT OF A PARALLEL CORPUS

Edyta Jurkiewicz-Rohrbacher

Doctoral dissertation, to be presented for public discussion with the permission of the Faculty of Arts of the University of Helsinki, in Auditorium XV, University Main Building, on the 13th of May, 2019 at 12 o'clock.

Proofreading: Kate Sotejeff-Wilson
Typesetting: LaTeX

ISBN 978-951-51-5157-5 (PB)
ISBN 978-951-51-5158-2 (PDF)

Unigrafia
Helsinki 2019

The Faculty of Arts uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

To my mother

Abstract

The objective of this contrastive research is (1) to determine the rules of correlation between the language-specific category Polish Verbal Aspect (PVA) and the elements of Finnish clause, whilst (2) re-examining the semantic scope of PVA, and (3) improving the definition of the cross-linguistically valid comparative concept of aspectuality.

The investigation is empirical, and based on 900 Polish-Finnish clauses compiled in the form of a bidirectional parallel corpus stratified in three samples according to text types.

The corpus is annotated on three levels, following the scalar model of temporality: the morphosyntactic and semantic clause-internal levels, as well as the clause-external level, including such elements as taxis and the quantificational-pragmatic context, temporally located (existentially quantified) situation, and generic or generalising interpretation (universal quantification).

The reasoning in the study is mostly inductive. In contrast to the previous studies on aspect, the work is organised bottom-up. The data is approached quantitatively, using state-of-the-art methods. First, the descriptive statistics of temporal markers in the corpus are discussed. Afterwards, the data is summarised in a statistical model and visualised in a hierarchical cluster structure. Particularly interesting correlations (e.g. tense-aspect or case-aspect) are further validated with the random-forests method.

The quantitative results yield a two-layered model of aspectuality, distinguishing between two levels: the outer, temporal-deictic level and the inner level related to the notion of change in time. Thus, the study confirms the validity of multilayered concepts of aspectuality as previously postulated.

As to language-specific results, PVA correlates with Polish and Finnish tenses within the outer, temporal-deictic layer. This interaction involves the third element – temporal quantification. The inner layer is realised in Finnish in the predicate-argument structure, and therefore, the Finnish argument case-marking is the closest correlate of PVA. Here the most important systematic opposition are between the lative and essive semantic cases (including Translative and Essive), and between the Total and Partitive type of object. The notion of change which is the semantically relevant factor is treated as gradable opposition (next to the traditionally used polar and equipollent oppositions), and therefore, the for-

mal comparison between PVA and Finnish differential object marking is possible within the scalar description.

The Finnish derivational valency modifiers (transtivisers and detransitivisers), however, do not seem to play any significant role in the marking of aspectual oppositions. Neither do lexical temporal expressions play much role here, as their generally low frequency does not deviate from the frequency of Polish expressions of that kind. In particular, the study shows that the measure adverbials in the object cases are quite infrequent in language use. Therefore, their contribution in expressing aspect is marginal.

Acknowledgements

A scientific study is the fruit of both the researcher's work and of the surrounding community. This book would not have been possible without supportive, understanding, inspiring and friendly people. I would like to thank them all.

First and foremost, I am thankful to my supervisors; Jouko Lindstedt, who guided me, helped straighten my thoughts and supported me through the whole writing process; Björn Hansen, who accepted me in the Institute of Slavic Studies at the University of Regensburg, and whose research seminars were always a platform for constructive discussions.

I would like to thank Mikhail Mikhailov and Ruprecht von Waldenfels for the preliminary examination of this book and useful comments, which I could benefit from in the final stages. I also feel very lucky to have had the book proofread by an amazing translator and editor, Kate Sotejeff-Wilson who was able to follow all three languages used in this study. Thank you for your patience and great job, Kate.

The manuscript could not have been written without a plan. I was lucky to meet Michael Oakes, who suggested improvements in the quantitative analysis, Laura Janda who gave me useful literature tips, Terhi Peltola who helped me in writing my first research plan, and Asia Richmond who improved my English.

Thanks to these good people, my work could start, but it would not have continued without support from various institutions. Thanks to the European Union I could work and study in several countries at the same time unhindered by many formalities. As to financial support, the University of Helsinki Research Foundation granted me my first scholarship; afterwards the Langnet and HELSLANG doctoral programmes for Language Studies and the University of Regensburg provided me salaried research positions and enabled participation in various summer schools and conferences.

Earlier versions of particular chapters were read and discussed by many people. First and foremost, by the members of the Department of Modern Languages at University of Helsinki and the Institute of Slavic Studies at the University of Regensburg. It was also a great pleasure to participate in PhD workshops within language programmes at the universities of Helsinki and Stockholm, winter night seminars and summer conferences of HELSLANG. Special thank is directed to the commentators, participants of this workshops, but also to Riho Grünthal, Gaïdig

Dubois, Heidi Niva – the organisers of HELSLANG meetings. These events really motivated me to writing.

I have been fortunate to discuss some problems with Jeremy Bradley, Dušica Filipović Đurđević, Elżbieta Kaczmarska, and Alexandr Rosen. Zrinka Kolaković was the driving force behind a few aspectological round tables with Volkmar Lehmann and Jurica Polančec.

I owe a debt of gratitude to good souls in Finland: the best colleague ever Johanna Virkkula, Hannu and Maria-Liisa Tommola, Linda Dejedarová, Turid Farbregd, Saarni Laitinen, Dusica Božović, Riikka Pulkkinen who gave me shelter from time to time and found time for good chats.

I would like to thank my friends with whom I could share my joys and sadness, and who I can always count on: my soul-sister and the best Finnish teacher I have ever had, Maria Sarhemaa, Ania Grzeszak, Kasia and Michał Ornatowscy, Ania Szymanska, Tobi Blessing, Beata and Tuomas Mäihäniemi, Iwona and Jari Kiuru.

I owe much to my Polish-German family: my siblings-in-law, nieces and nephews; my parents-in-law who supported my family and the writing even in the gloomy days of my father-in-law's illness. My grandmother Inez Wiatr has always been an example of a brilliant academic. I thank the best parents in the world, Monika and Zbyszek Jurkiewicz, for believing in me and being tolerant of my most weird ideas. I am lucky to have such great brother Kuba; your little sister learned so much from you. The biggest kisses are for my daughter Hania. Thank you for bringing so much joy and brightness into my life, preventing me from sitting too long in front of the computer, and asking fundamental questions about the nature of world. Finally, I am greatly indebted to Stefan, my patient, understanding, and caring husband. Thank you for encouraging me to do things I would never dare without you and for cheering me with your music.

Augsburg, 03.04.2019

Contents

List of Figures

List of Tables

List of abbreviations

1 First person

2 Second person

3 Third person

ABL Ablative

ACC Accusative

ACCUM Accumulative

ADE Adessive

ADJ Adjective

ADV Adverb

AG Agent Participle (Finnish)

ALL Allative

AUX Auxiliary

COM Comitative

COND Conditional

COP Copula

DAT Dative

DISTR Distributive

DOM Differential Object Marking

DELIM Delimitative

ELA Elative

ESS Essive

F Feminine

FOC Focus particle

FREQ Frequentative

GEN Genitive

HAB Habitual

ILL Illative

IMP Imperative

IMPS Impersonal

INCH Inchoative

INE Inessive

INF Infinitive

INF1 The First Finnish infinitive

INF2 The Second Finnish infinitive

INF3 The Third Finnish Infinitive

INS Instrumental

INSTR Instructive

IPFV Imperfective

ISC Involuntary State Constructions

LPTCP L-Participle (Polish)

LAT Lative

LOC Locative

M Masculine

MOM Momentanuous

N Noun

NEG Negative particle

NEU Neuter

NVIR Non-virile

NOM Nominative

PSTAP -nUt Participle (Finnish Past Active Participle)

ADJPAS Passive adjectival participle (Polish)

OBJ Object

OSMA Measure adverbial in the object cases

PAR Partitive

PASS Passive

PERDUR Perdurative

PFV Perfective

PL Plural

POSS Possesive suffix

PST Past (Polish)

PVA Polish Verbal Aspect

REFL Reflexive

SEMEL Semelfactive

SG Singular

~~IPFV~~ Imperfective marker whose meaning is cancelled by another marker

~~PFV~~ Perfective marker whose meaning is cancelled by another marker

SPST Simple Past (Finnish)

SUFF Suffix

TBS To-be-spoken

TOT Total Case

TR Referential Unit of Time

TRANS Translative

TSIT Time of Situation

TU Time of Utterance

PSTPP -tU Participle (Finnish Past Passive Participle)

V Verb

PP -vA Participle (Finnish Present Active Participle)

VIR Virile

Chapter 1

Introduction

1.1 Topic of the study

In the present study, I depart from the language-specific category Polish Verbal Aspect (henceforth: PVA), an obligatory verbal classifier based on a binary opposition Perfective (PFV) – Imperfective (IPFV) illustrated below:

(1) a. *Pis-a-ł-e-m* *list.*
write-IPFV-PST-M-1SG letter.ACC
'I wrote/was writing a/the letter.'

b. *Na-pis-a-ł-e-m* *list.*
on.PFV-write-~~IPFV~~-PST-M-1SG letter.ACC
'I wrote a/the letter.'

(2) a. *Pociąg za-trzym-ywa-ł* *się* *na stacji.*
train behind.~~PFV~~-hold-IPFV-PST REFL on station.LOC
'The train was about to stop at the station.'

b. *Pociąg za-trzym-a-ł* *się* *na stacji.*
train behind.PFV-hold-~~IPFV~~-PST REFL on station.LOC
'The train stopped at the station.'

The situations in (1b) and (2b) are interpreted as performed within a discrete

(impossible to quantise) unit of time[1], but no such constraint applies to the situations in (1a) and (2a).

In many languages – including Finnish, the other language studied here – no explicit, regular tools for marking such oppositions can be identified. In Finnish, the semantic notion of aspectuality became part of the non-normative description only recently (Hakulinen et al. 2004). Thus, although the need for describing the phenomenon of aspectuality has been acknowledged, no domain of it seems to be clearly grammaticalised in Finnish. Nevertheless, the meanings presented in (1) and (2) can be expressed in Finnish too, though on the level of the clause, not the verb:

(3) a. *Kirjoit-i-n* *kirje-ttä.*
write-SPST-1SG letter-PAR
'I wrote/was writing a/the letter.'

b. *Kirjoit-i-n* *kirjee-n.*
write-SPST-1SG letter-GEN
'I wrote the letter.'

(4) a. *Juna ol-i* *pysä-ht-y-mä-ssä* *asema-lle.*
train AUX-SPST stop-MOM-REFL-INF3-INE station-ALL
'The train was about to stop at the station.'

b. *Juna pysä-ht-y-i* *asema-lle.*
train stop-MOM-REFL-SPST station-ALL
'The train stopped at the station.'

It is important to observe that situations referred to in (1) and (3), and in (2) and (4) differ in predicate-argument structure. (1) and (3) describe a quantitative change which applies to the patient. In both sentences the minimum quantity of the patient might be equal to zero (the patient does not exist yet), but they differ in respect to the final quantity of the patient reached in the situation. In (3a) the quantity is greater than zero but unspecified. In (3b) the quantity of the end-state reaches the maximum value one. The difference can be concluded from the difference in case marking of the Finnish direct object.

[1]This does not mean that the time needed to perform these situations is not measurable, but the scope of possible expressions is limited, as I explain in Sections 2.4 and 3.8.3.

Examples (2) and (4) describe a movement a towards goal. In these cases, the predicate-argument structure describes the position of the mover in reference with the end-point of the path. This time the difference is visible in different forms of the predicate.

Thus, the notion encoded in verb structure in Polish is not realised in Finnish by one grammatical category, but it has various means of expression.

1.2 Goals of the study

The current study aims to examine the possible aspectual markers in Finnish departing from the contrast with the well-identifiable category of PVA. In particular, I am interested whether any particular types of predicate-argument structure in Finnish are important, and if so, which arguments are relevant. Secondly, although PVA opposition has clear formal tools of expression, its semantics remains fuzzy in comparison to the scope of functions and temporal interactions observed in previous studies (cf. Bartnicka et al. 2004; Holvoet 1989; Laskowski 1998b; Śmiech 1971). Thus, the second goal of the study is re-examining the semantics of PVA. The comparison of realisation of aspectual categories in Finnish and Polish leads to the revision of the cross-linguistic definition of aspect, or better formulated, the cross-linguistically valid comparative semantic concept of aspectuality.

PVA is often said to contribute primarily to "different ways of viewing the internal temporal constituency of a situation" Comrie (1976: 3), that is to the subdomain of temporality (or a domain overlapping with temporality), called *aspectuality*. In that respect, PVA is compared to such categories as the English Progressive, or Romance Simple Past and Imperfect.

On the other hand, some scholars (Bertinetto & Delfitto 2000; Dahl 1985) conclude that Slavic-style aspect, to which PVA belongs, is not the typical perfective–imperfective aspect. One reason, given by Dahl (1985) is the fact that PFV is not limited only to the past temporal reference. Additionally, although the formal opposition PFV – IPFV can be identified in all Slavic languages, previous studies (Dickey 2000; Gvozdanović 2012; Stunová 1993) have shown that the functional-semantic scope of Slavic aspect varies across members of the group. In particular, Stunová (1993: 193) concludes major differences between verbal aspect in Russian and in Czech, as the Russian verbal aspect serves the global discourse strategy, while the Czech aspect is involved in marking the inner temporal structure of each situation separately. The research suggests that comparisons with Polish would be interesting here, as Polish verbal aspect is known to behave in some

respects like the Russian aspect and in some like the Czech one. In the present work, I intend to defend the claim that the scope of functions of PVA is broader than only specifying "*the internal* temporal constituency".

1.3 Aspect in Finnish from the Finnish-Slavic contrastive perspective

As stated above, Finnish does not have any obvious grammatical markers corresponding to Slavic aspect opposition. However, the questionnaire based study of Dahl (1985) showed that from the typological perspective the Finnish diffential object marking (henceforth DOM, see Sections 4.2.2) is close to the same phenomenon as Slavic style aspect. Therefore it comes as no surprise that studies with the Finnish object in focus approach its case marking in terms of aspectuality (Askonen 2001; Heinämäki 1984, 1994; Larjavaara 2007).[2] I elaborate on this category, which is central for Finnish aspectology in Section 4.3.

Also Finnish-Slavic contrastive studies focus mainly on the relation between the aspect and DOM (Tommola 1986; Zmrzlíková 2009). Both works discuss particular features of Finnish grammar in the light of contexts where particular values of aspect in Russian or Czech are used. Both works refer to literary data, but only Tommola (1986) includes quantitative summaries.

The main focus is naturally placed on DOM. Tommola (1986: vii) states that the Total object contains the semantic feature of what he calls *resultativity*: "specificness of the object concept and specificness of the end state resulted from the action" – which is proximate to the Russian verbal aspect. However, Tommola characterises Russian verbal aspect as governed by two features: *boundedness* (Rus. *predel'nost'*, the existence of a bound limiting the situation) and *totality* (Rus. *celostnost'*, the non-divisibility of situation's structure). Resultative situations are total. However, in Russian the non-total, but bounded situation may be expressed with both aspects, thus also with PFV, although in Finnish such situations must be marked with the Partitive object.

Both authors discuss the role the measure adverbials in the object cases, and verbal affixes which modify the temporal structure of situation. (Tommola 1986)

[2] Additionally, aspect is often approached in Finnish linguistics from the cognitive perspective (Huumo 2006; Nurminen 2011, 2014, 2015, 2017; Sivonen 2007). I leave the cognitive direction without a comment, because the starting point for this inductive research is formal grammatical category, and not the cognitive definition of the term *aspect*.

also analyses aspectuality in the context of tenses and the lative-essive distinction (see Section 4.2.2).[3]

The expression of aspect appears in Finnish as clausal phenomenon and its realisation is syntactically motivated. Kangasmaa-Minn (1984) explains that with the fact that verbs in Uralic are not as rich in information as, for example, in Slavic languages. Therefore, aspect is encoded in nominal dependents of the predicate, as they carry most information in the sentence. Similarly to Tommola (1986), she points at the importance of the lative-essive distinction encoded in the system of Finnish cases.

Biskupska (2018) compares the verb derivation systems in Polish and Finnish. In Polish deverbal lexemes are often derivied by means of spatial prefixes, which directly influence the aspectual value assigned to the lexeme (see Section 3.5.5). Additionally, the reflexive marker *się* oscillates between the status of clitic and affix (see Section 3.3), but its relation to PVA is unclear. In Finnish, the two most common groups of derivational affixes concern the change in number of arguments (see Section 4.2.1). Biskupska concludes that Polish and Finnish derivatives differ with respect to their semantic scope, in particular as to the notion of the change of state and aspect, which in Polish are more salient thanks to prefixation.

1.4 Data and methodology

1.4.1 Bottom-up approach to contrastive studies on aspect

As indicated above, PVA does not have straightforward counterparts in Finnish. Additionally, previous reseach suggests that identifying these correlates requires considering possibly broad context, minimally the unit of clause. In my view, this requires turning away from the traditional deductive reasoning used in the previous studies on aspect, and following the more agnostic, inductive approach.

Consequently, the present work is empirical and organised bottom-up. No hypothesis is assumed a priori and tested *against* the data, but the conclusion is drawn directly *from* the data and reflected in the light of existing theories. In order to achieve this goal, I use statistical exploratory methods. I base my findings on the empirical data stored in the form of parallel corpus, that is, original texts aligned to their translations.

[3] I return to these studies in Chapter 8 in order to contrast their quantitative results with my own.

1.4.2 The distributional hypothesis

The main methodological assumption of this study arises from the *distributional hypothesis* related to the work of Harris (1954), namely, that linguistic elements with similar distribution in texts belong to the same semantic-functional category (cf. Sahlgren 2008: 33). In other words, when two linguistic elements, for example, e_1 *nice* and e_2 *beautiful* occur regularly with another linguistic element e_3 *girl*, one may assume that e_1 and e_2 belong to the same linguistic class (in that case for example to the class of adjectives).

The DISTRIBUTION OF LINGUISTIC ELEMENT "is a sum of all environments in which a linguistic element appears, and an ENVIRONMENT of a linguistic element is an array of its co-occurrents, i.e. the other elements (...) with which an element occurs to yield an utterance" (Harris 1954: 146).[4]

1.4.3 Data

The studied sample covers indicative, affirmative clauses in Polish and Finnish which contain simple predicate forms, that is, predicates consisting of one finite form. Thus, the study excludes infinitival complements and participle clauses. The clauses originate from the parallel Finnish-Polish texts (originals and their translations) obtained from various written sources. The corpus (see Chapter 5) is bidirectional, so both Polish and Finnish originals are included in equal proportions. The final data set, which consists of 900 parallel clauses, can be stratified into three subsamples according to text type: literary-narrative, informative and to-be-spoken. Literary-narrative texts are obtained from fictional texts, informative sample covers news and essays. The to-be-spoken type includes play scripts, film subtitles, and dialogues extracted from literary texts. The text type stratification is motivated by significantly different tense-aspect discourse structures in the chosen samples, as shown in Section 5.6.

1.4.4 Methods

The annotation of corpus requires taking into accont the temporal systems of both languages in question. The lack of suitable framework allowing for comparisons between different temporal systems is one reason why little cross-linguistic work has been done (Dahl 2000: 3). The functional (Bondarko 1991: 64-94) and

[4] Harris (1954) includes in the definition of environment also the particular position of elements. In the current approach, I omit this constraint.

cognitive-functional (Bartnicka et al. 2004; Dickey 2000; Lehmann 2009) models can be applied to cross-Slavic comparisons, but I see their weakness in assuming a limited (therefore not necessarily exhaustive) set of functions where the difference between PFV and IPFV is relevant. Since aspect is undeniably related to temporality, and time is usually subject to measurement, one handy approach to examining temporality is scalarity (see Section 2.6). Therefore the temporal systems of Polish and Finnish (see Chapters 3 and 4), within which aspect can be characterised, are described in the present work according to one scalar-temporal model (see Chapter 2). Afterwards, the corpus is additionally annotated for clause-internal morphological, semantic, syntactic and clause external features such as text type or temporal quantification.

The data set is analysed with a set of advanced quantitative methods. The linguistic features are preliminarily explored for their frequency and distribution. This leads to two conclusions: 1.some features are rather infrequent and/or sparsely distributed, 2.the system of interdependencies is so complicated that it cannot be summarised with simple significance-testing methods.

Therefore, the most frequently occurring features are further summarised in statistical models. First, the similarity between semantic, grammatical and lexical features is explored with the notions of similarity and distance upon which a hierarchical cluster tree is built to show the data structure. The validity of the most informative clusters is further tested with *random forests* (Breiman 2001) with which I try to find out whether the value of PVA can be predicted directly from the most frequent Finnish temporal features.

The random-forests model, cluster analysis and descriptive statistics of the data are used to draw the final conclusions about PVA and its correlates in Finnish.

1.5 Organisation of the book

Following this introductory chapter, in Chapter 2, I construct the scalar-temporal model within which Finnish and Polish are contrasted with respect to PVA. Having described different parameters of the temporal domain, I introduce the concept of scale and apply it to temporality. Chapters 3 and 4 are devoted to language-specific characteristics of temporal systems acknowledged in the literature. In the case of Polish the interactions between PVA and particular temporal domains are discussed within the scalar model. Chapter 5 describes the structure of the Polish-Finnish parallel corpus, while Chapter 6 is devoted to the implementation of temporal description in the form of annotation in the parallel corpus.

Chapter 7 presents the quantitative results of the investigation. I discuss the results in the context of empirical questions formulated in Section 7.1. In the first part, the descriptive statistics of temporal markers in the corpus are discussed. Afterwards, the data is summarised with two statistical methods – hierarchical agglomerative clustering of a distance matrix calculated based on a simple match coefficient, and recursive binary partitioning – and visualised in the form of tree structures.

In Chapter 8, I evaluate the results of the quantitative analysis in the light of the previous theoretical claims about PVA known from the literature. Secondly, the Finnish correlates of PVA are reconsidered as the functional correlates of the grammatical category in question. Finally, I comment on the contribution this study makes to describing the cross-linguistically valid comparative semantic concept of aspectuality. The chapter closes with discussion of the limitations of the study and further research possibilities.

Supplementary information, such as text sources or frequency lists are given in the appendices.

1.6 Note on adopted style conventions

Regarding layout, I mostly follow the most recent edition of the *Unified style sheet for linguistics* of the Committee of Editors of Linguistic Journals, summarised in the *Generic Leipzig style rules*.[5]

The reader will find the list of glosses and abbreviations in the beginning of the book. Since *The Leipzig glossing rules*[6] contain only very basic glosses, I had to extend them by the number of glosses necessary to show the structure of Polish and Finnish grammatical categories.

The glosses of examples are enumerated and always contain three lines: the example line, glossing line and translation line. If an example is taken from my own corpus (to which I refer as corpus_2, see Section 5.1), the ID of the sentence in corpus_2 is provided together with the translation, so the exact source can be retrieved from the list of corpus sources given in Appendix B. In other cases, the exact source is given, for web sources in the form of a hyperlink to the website, which can be found in the appropriate footnote.

Whenever an example contains the Finnish and the Polish version of a clause, first the original clause is given (in *a*) followed by the translated clause (in *b*).

[5] https://www.eva.mpg.de/lingua/pdf/GenericStyleRules.pdf
[6] https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf

The key terms with their definitions are introduced with SMALL CAPS, while other, metalinguistic terms are written with *italics*. This layout is also used for all object-language forms together with translation in single quotation marks. The language-specific categories of tense, aspect and case are described using the definite article *the* and capitalised, which I hope, will help the reader.

Chapter 2

The relation between temporality and scales

2.1 Introductory remarks on temporality

Since the current study is mainly inductive, I avoid making many theory-based assumptions about the nature of aspect. Instead, I place it in the possibly broad context of temporality and later try to show how PVA interacts within the domain of temporality as stated in Section 1.4.4.

Following Lindstedt (2001: 768), a SITUATION stands "for anything that a sentence denotes, or an utterance refers to – an event or state, for instance". Situations are assigned to spacetimes and each natural language is capable of expressing the spatiotemporal properties of a situation. The properties of the temporal course of a situation (TSIT) form a semantic-functional field which I call TEMPORALITY. Temporality concerns three main questions:

1. 'when?' (or 'where in time')
2. 'for how long?'
3. 'how many times?'

The first question refers to assigning TSIT to the referential temporal unit (TR) belonging to the time axis. Such an operation is called TEMPORAL LOCALISING (see Section 2.2). DURATIVE TEMPORALISATION means measuring out the length of TR to which TSIT is assigned,[1] and it helps answer the second question

[1] In the present work I do not study the nature of time, but rather its perception as reflected in

(see Section 2.3). Finally, answering question three, language tools from the field of temporality may assign TSIT to more than one TR (see Section 2.5).

In the present study, I investigate declarative, affirmative clauses, which appear in the corpus. Thus, the studied material consists of utterances, and not of system sentences which could be ambiguous with respect to some of their temporal parameters, in particular to temporal quantification, as explained in Section 2.5.2. Thus, the situations discussed here are tokens of situations appearing in utterances, and not situation types appearing in contextless clauses. Modality, which normally is a property of a clause, is excluded from this study.[2]

Pure temporality is a spatio-temporal description with only one salient parameter. However, there is nothing uncommon in merging spatial and temporal characteristics in one unit, as in the expression *Ancient Greece* in the sentence below:

(5) *In Ancient Greece people wore togas.*

In this chapter I give some examples of the most widespread and most basic ways of expressing temporality in language. I return to the spatial parameter later, while analysing the data. Although the three temporal subdomains are relatively easy to distinguish, languages neither keep them separate nor develop exclusively grammatical or lexical tools of expression. Instead, they form complex grids of morphological, lexical and syntactic relations.

2.2 Temporal localising

2.2.1 Temporal adverbials

A situation can easily be localised in time using temporal adverbial expressions. Two elements possible in adverbial expressions must be distinguished: the name of the unit of time, for example, as a date: *January the first*, and the type of relation between the TSIT and the unit of time (e.g. ***on*** *the January the first*).

As pointed out by Laskowski (2003) there are two types of adverbial temporal expression. After Haspelmath (1997: 25), he calls the first group *canonical time periods* such as calendar dates, names of parts of the year, or time units such as an

natural languages. Time may be experienced by language users in units measurable on continuous or discrete scales. The expression unit of time is neutral and can mean any part of the time axis of any properties.

[2]Nevertheless, I do not resign from the study of future temporal reference.

hour or second. Names of events (e.g. *breakfast, the Second World War*) can also be classified in this category. The second type consists of *subjective time periods* such as *a moment*.

Some temporal expressions are deictic, as they relate TSIT to TR indirectly, in relation to TIME OF UTTERANCE (TU) such as adverbs – *yesterday*. TEMPORAL PARTICLES[3] (like *yet*, *already* or *still*) compare TSIT against some mental standard, for example, speaker's expectation of how long the situation shall last, or when it shall terminate.

2.2.2 Deixis

The grammatical category related to temporal localising is TENSE.[4] The three primary distinctions of the *absolute tenses* (Comrie 1985: 36) refer to contrast caused by different relations to the TU:

- TSIT's overlap or coincidence with TUis represented in English by the Present tense, as in the sentence:

  (6) ***I am dancing***.

- TSIT anterior to TU is represented in English by the Past tense as in the sentence:

  (7) ***I was dancing***.

- TSIT posterior to TU is represented in English by the auxiliary Future tense as in the sentence:

  (8) ***I will be dancing***.

[3]Sometimes called aspectual particles.

[4]Some problems related to accurate definition of tense for the purpose of contrastive or typological studies are given by Haspelmath (1997: 6) who examines the definition of Comrie Comrie (1985: 9): "grammaticalised expression of location in time". Haspelmath points out several weaknesses of this definition in comparison to what Comrie really aims to describe in his study. First, the original definition covers morphemes appearing in nominal phrases such as the preposition 'in' in *in the spring*, the Finnish Adessive in *kevää-llä* or the Polish Instrumental in *wiosn-ą*, which have the same meaning as the English expression. However, including the verbal constraint in the definition would exclude the validity of the definition for languages with the nominal tense. Since neither Polish nor Finnish have the latter category, I consider tense a purely verbal category.

2.2.3 Taxis

Situation S_1 does not need to be temporally localised in the relation to TU, if it is possible to establish its relation to situation S_2 for which TSIT has already been localised. This type of temporal localising is called TAXIS (Jakobson 1957[1971]).[5] An example of taxis marked with tense is shown below:

(9) *And when he had said this he disappeared.*[6]

In (9), the tense used in the main sentence clause defines temporal localising prior to TU, while the Pluperfect appearing in the subordinate clause serves two functions. First, it also refers to a situation anterior to TU (*absolute* temporal localising), and, second, it shows that the situation in dependent clause is anterior to the situation referred in the main clause (temporal localising *relative* to the temporal localising of TSIT of the main clause). The Pluperfect could thus be considered an ABSOLUTE-RELATIVE tense.

While in some languages taxical relations are expressed with tenses (e.g. in English, as shown above), Maslov (1978: 8-9) notices that in many languages taxis cannot be a separate grammatical category, but is included in a combination of tense and aspect. In such cases, deictic tenses are used in the taxical function, that is, as relative tenses. According to Maslov, expressing taxis is one of the most important function of Slavic Verbal Aspect (see Section 3.8.2).

Taxis can be also expressed lexically, by means of ordering expressions such as *later*, *afterwards*, *at the same time*.

2.2.4 Relativity of temporal localising

The same temporal properties may be expressed by different means:

(10) a. *I bought the car on **January the second**.*
b. *I bought the car **a week ago**.*
c. ***After you had told me to stop biking**, I bought the car.*

[5]The understanding of the term taxis. In the present work, the scope of taxis is limited to two clauses belonging to the same sentence or connected with a lexical (e.g. anaphoric) marker. Taxis is examined in terms of the TSIT of the dependent clause, or in the case of two clauses connected with a coordinate conjunction, TSIT of the clause appearing later in the linear order is relative to the TSIT of the first one.

[6]`https://www.thedivinemercy.org/news/Why-Do-We-Call-Mary-Mother-Of-Mercy-3369`

In a particular context, in all three sentences the verb *to buy* can refer to the same situation, but from different perspectives: in terms of objective date (10a), of deictic distance (10b), and as posterior to some other situation (10c).

Example (10) shows that one temporal dimension can be expressed by various linguistic means. Languages have different inventories of categories with which they express the same temporal dimensions. The category lacking in some language can be substituted by a combination of some other categories (either lexical or grammatical). For example, although Mandarin Chinese is considered as tenseless language (cf. Lin 2012), the language enables temporal localising in other ways, such as aspectual information, temporal adverbials or discourse anaphora.

2.3 Durative temporalisation

Durative temporalisation is more complicated than temporal localising. This is because the perception of lasting in time is subject to personal evaluation.

While lasting itself can be measured and expressed quite objectively in adverbial expressions specifying lasting in terms of length of time (***for*** *two hours*), only its left or right boundary (***from*** *dusk* ***till*** *dawn*), or naming a unit of defined duration (*whole week*), other expressions are less precise and more subject to personal impression as they either refer to lasting qualitatively (*short*, *long*) or are imprecise per se (*a while*, *a moment*).

2.4 Temporal localising, durative temporalisation or something else

Three types of expression which combine the marker of relation with a temporal interval are not easy to classify either as tools of temporal localising or durative temporalisation.

The former include temporal frames within which the situation took or will have taken place (*in two hours*, *during last two months*, *within the next hour*):

(11) *Look at your dad, he died in three days.* (BNC 2007: KBB 6994)

Such expressions may be called FRAME ADVERBIALS (in Stawnicka 2007: 129 *durative frame adverbials*) as they provide an interval within which the TR to which TSIT is assigned is included but it does not necessarily fill this interval

entirely. In other words, the TR to which TSIT is assigned in (11) is not equal to the whole interval of *three days*. I elaborate on this in Sections 3.8.3 and 4.5.3).

The other two types do not apply to TSIT referred to by the main verb but to some time posterior to TSIT:

(12) *Further talks are scheduled* ***for 16 October***. (BNC 2007: A30 666)

(13) *But Moon still hopes to be fit for the Wales squad that leaves* ***for a week of warm-weather training in Lanzarote*** *on Tuesday.* (BNC 2007: CBG 3455)

Both types can be characterised as *prospective*, but, as mentioned above, they do not refer to TSIT of S presented in the sentence. In (12), *for 16 October* is a marker of TR when *further talks* should take place. In (13), *for a week* means 'in order to spend somewhere as much time as'. Therefore, it should be treated as a marker of degree.

Date does not need to fulfil a temporal function:

(14) *Mutually agree on* ***a suitable time***.[7]

a suitable time can be considered as topic or content, as in the phrase *agree on meeting*.

2.5 Pluractionality

2.5.1 Parameters of pluractionality

The term PLURACTIONALITY was introduced by Newman (1980). In the present work, I understand pluractionality in the narrow sense, as the capacity of assigning TSIT to distinct TRs.

As explained below pluractionality has three parameters:

- temporal quantification (specificity of assignment to TR)
- quantification over referents (individual-level versus stage-level or kind-referring versus non-kind-referring)
- type specifying the frequency of repetition

I will explain those parameters below.

[7] `https://www.wikihow.com/Do-a-Handover-in-an-Office`

2.5.2 Temporal quantifiers

There are two main types of temporal quantifier: existential and universal. In the first type, occurrences of situations are described, and TSIT is assignable to TR, but the degree of certainty about TR of individual occurrences varies:

(15) a. specific
Yesterday, I ate lobster for dinner.

b. non-specific
I have been to Warsaw.

Specific and non-specific sentences represent occurrences of situations. They refer to situations anchored to at least one unit belonging to some timespace. In the case of a specific occurrence, it is possible to logically evaluate the following expression: 'There exist(s) a particular moment(s) X when S happens' as true or false. In non-specific cases, the expression must be transformed as follows: 'There exist(s) *some* moment(s) X when S happens.' Thus, in specific temporal quantification, the TR to which TSIT is assigned is known and probably relevant for the utterance, while in non-specific temporal quantification, the speaker does not know TR, or does not consider it relevant for the utterance.

In universal quantification, TSIT cannot be assigned to any particular TR:

(16) a. *The lion has a bushy tail.* (in the sense: 'Each lion's tail is bushy.')

b. *Four is an even number.* (in the sense: 'Being an even number is a property of number four.')

(17) a. *Maria dances nicely.* (in the sense: 'If it happens that Maria is dancing, she is doing it nicely.')

b. *The diplodocus ate leaves.* (in the sense: 'If animals representing the species diplodocus ate something, those were leaves.')

Sentences quantified universally are continuously valid, but they do not describe occurrences of situations. Nevertheless, the sentences in (16) differ from those in (17). The former are STATEMENTS used to formulate some general, omnitemporal laws; the latter reflect PATTERNS of situations.

2.5.3 Quantification over referents

The possibility of assigning TSIT to TR depends on the type of referents, which may be either:[8]

- individual or generic[9]
- real or abstract

The difference between individual and generic types of reference has much in common with existentially and universally quantified situations, that is, between occurrences and patterns of occurrences. In this case, the distinction is between an individual and the class to which the individual belongs. As shown in (17), universally quantified sentences may apply to both individual (17a) and generic referents (17b).

Some referents are seemingly existentially quantified as below:

(18) ***Pohjolan asukkaatkin*** *tutustuivat* *1800-luvulta*
North inhabitant.PL.too acquaint.SPST.3PL 1800.century.ABL
alkaen ***kaupunkimaiseen elämänmuotoon***
starting city.like.ILL life.style.ILL
'**The inhabitants of the North** also became familiar with **the city life-style** starting from the nineteenth century.' (S63)

(19) ***Tavallinen kansa*** *kukaties tuli* *tekemisiin* ***tiedostetun***
ordinary people perhaps come.SPST doing.PL.ILL conscious.GEN
ujouden *kanssa vasta sitten kun* ***koululaitos ja muut***
shyness.GEN with only then when school and other.PL
modernin yhteiskunnan instituutiot ja tavat *1800-luvulla*
modern.GEN society.GEN institution.PL and custom.PL 1800.century
tunkeutuivat *sen* *keskuuteen.*
push.way.SPST.3PL it.GEN among.ILL
'**The ordinary people** started to cope with **conscious shyness** probably only when **school and other modern society's institutions and customs** pushed in in the nineteenth century.' (S77)

[8]The number of referents should also be accounted for: this usually influences the situation-internal pluractionality.

[9]Or: individual or stage level Shluinsky (2009); non-kind-referring or kind-referring (Krifka et al. 1995).

Nevertheless, it is hard to assign the TSIT of situations to any specific TRs even though a temporal localising expression (*1800-luvulla*) 'in the nineteenth century' or durative temporalisation (*1800-luvulta alkaen*) 'starting from the nineteenth century' are given.

The reason for that is the abstract character of the bolded referents in the sentences, which support the case against the existential temporal quantification of the sentence. The more abstract the referent, the more probable it is that the TSIT cannot be assigned to any particular TR.

Nonetheless, abstract entities can also appear in specific situations:

(20) ***The European Parliament*** *voted today on Mr Casaca's report on discharge in respect of the implementation of the budget of the European Parliament for the 2007 budgetary year.*[10]

2.5.4 Specifying type of frequency

The way frequency is specified in the sentence complements the distinction between universal and existentially quantified situations and fills in the continuum between TSITs of existentially quantified situations assigned to one TRs and TSITs of universal situations assignable to an infinite or unknown number of TRs. Therefore, it makes sense to distinguish between at least three basic types of frequency marking:

- summaric (e.g. *seven times*, *twice*)
- specific cycle type (e.g. *daily, every two hours*)
- unspecific cycle type (e.g. *often*, *rarely*, *continuously*, *usually*, *always*)

Not only adverbial expressions may be used to specify the frequency, but languages use particular grammatical forms or constructions to express, for example, events of specific (21) or unspecific cycle (22) type, such as English *used to*:

(21) *Hanna used to eat cake in this pastry shop every Friday.*

(22) *Hanna used to wear blue shoes.*

[10] `http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+CRE+20090423+ITEMS+DOC+XML+V0//EN\&language=EN\#creitem7`

2.6 A scalar approach to temporality

2.6.1 Measurement and scales

The above discussion shows that temporality is a complex field involving several subfields. The means of expression of particular subfields constitute a broad spectrum, which makes comparing the features of two language systems more challenging. One must consider not only the language-specific, grammatical(ised) categories but also the potential interactions. Following the argumentation from Section 1.4.4, I explain now how scale can be used as framework allowing for comparisons of different temporal systems.

MEASUREMENT is "a *procedure* by which one assigns a given *individual* in a *domain* to exactly one *category* in a collection of categories" (Hardegree 2001b: 11). The three purposes of measurement are: comparison, classification, and communication (Hardegree 2001b: 5).

A distinction needs to be drawn between DIRECT and INDIRECT comparison. Direct comparison requires only the two compared objects, while indirect comparison requires the third STANDARD object against which an individual is compared. Standards are conventions necessary for communication.

The collection of categories to which one assigns individuals, in other words classifies them, is called SCALE. Scales are often artificially introduced for measuring such properties as weight, time, or temperature, but one should still bear in mind that our language is simply a scale providing collections of categories called words to which we assign real-life objects. Words and grammatical categories are therefore scales that I could call mental standards. The word 'white' represents the mental standard of a certain light frequency spectrum, while the category of tense is a scale on which situations are typically assigned as posterior, parallel, and anterior to some particular reference point – the standard object of comparison.

2.6.2 Types of scale

Stevens (1946) distinguishes four basic types of scale:

1. **nominal** scale – a collection of categories to which objects are assigned without any quantitative meaning.The extreme types of nominal scale are a universal scale with only one category, and an identity scale where each is uniquely classified. An example of a nominal scale is the category of gender (masculine, feminine, neuter).

2. **ordinal** scale – a collection of meaningfully ranked categories. Differences between particular categories can be compared only qualitatively. An example of an ordinal scale is the Likert scale, often used in questionnaires (e.g. strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).

3. **interval** scale – the differences between particular values of the scale (categories) can be compared quantitatively, but the values cannot be compared quantitatively. The Celsius scale is a well-known example of an interval scale. Although in the case of changes from 0°C to 10°C and 10°C to 20°C one can say the interval is 10°C, 10°C is not half as warm as 20°C.

4. **ratio** (additive) scale – both the differences between particular values of the scale and the values themselves can be compared quantitatively. For example, the height of a human being can be measured on the ratio scale.

For values which can be compared quantitatively, it is also possible to draw further distinctions:

- **discrete scale** – defines a finite or countable set from which the values of variable can be taken,

- **continuous scale** – a variable measured on a continuous scale can take any value between any two other values.

Often a variable can be measured on both scales and the continuous scale can be DISCRETISED by being divided into intervals which are countable.

2.6.3 Temporality in terms of scales

Given what has been said about temporality so far, I attempt to describe the scalar properties of the temporal features I have already mentioned.

The most relevant scale in terms of temporal localising is the ordinal scale, because the main function of temporal localising is ordering temporal units, TRs, linearly according to the ordinal scale of time. TR itself can be measurable on a discrete or continuous scale (see Section 3.7).

Durative temporalisation, in contrast, deals mainly with interval and ratio scales, because its main function is measuring the distance between two points which define lasting, or measuring the length of duration which can be compared

against other durations. Durative temporalisation is possible only in the case of TR which is measured on a continuous scale or discretised.

Finally, pluractionality (in particular the type of frequency) operates on a ratio scale, as TRs to which TSIT is assigned are counted. In this case, the scale itself can be discrete or continuous. Thus, summaric types of frequency are discrete, while when defining the type of frequency according to the cycle, both types of scale may be used.

.

2.6.4 Verbal arguments and measurement

The temporal properties of a situation are often specified by verbs, but languages map them also to nominal dependents. This notion is particularly important in the Finnish temporal system. Scholars consider the case of the direct object to be for the central marker of temporal progress (see Section 4.3). This is related to the fact that the *whole-part* concept is fundamental to human capacity to grasp the world (Hardegree 2001a: 3). Both events and objects are conceptualised as consisting of parts: cooking is a set of ordered procedures, machines are made of screws, plates, cables. Assuming that a situation involves some mediated material, this material has parts: a book consists of chapters, a dance is made of steps and figures.

The contribution of nominal phrases to temporality has been aconcern of formal semanticists such as Dowty (1991), Krifka (1992), Tenny (1994), and Filip (1993) who all conclude that some types of argument encoded mainly in a syntactic object are capable of "measuring out" the situation. Dowty (1991) calls this phenomenon *theme-to-event homomorphism*. For example, in the situation *eating an apple* the portion of eaten apple correlates directly with the progress of the situation, which terminates with a disappearance of the apple. Therefore, a temporal property of the situation can be mapped to the syntactic object, in this case the apple. This is not possible in the case of a situation of *liking Mary*, because the time of *liking Mary* applies always to the whole, indivisible unit. It cannot be assigned to parts of Mary (one cannot say that at the time t_1 he liked the hand of Mary, at t_2 the hand *and* the shoulder until at t_3 he liked *the whole* Mary).

The progress which is mapped to an argument can apply to the change in quantity or quality (all or part of) of the argument. It is represented in the above situation *eating an apple* (quantitative change) or *peeling an apple* (qualitative change). The homomorphism which is mapped directly to the change of argument I call ARGUMENT SCALE.

However, the change does not need to affect the argument to which the temporal progress is mapped physically. This is prototypically the case of movement-related situations, such as *putting an apple on the table*. In that case, mapping is the result of the change of position between the mover and the argument which marks the PATH SCALE. The position of the mover in relation to the path is homomorphic to the temporal progress.

Finally, in situations such as *reading a book*, *playing a song* the argument does not change, but is a medium onto which some sort of mapping is possible. As one cannot say a priori whether this type of homomorphism is more similar to the argument scale or to the path scale, I call this type INHERENT SCALE homomorphism.

The two types of scale – argument and path – are usually distinguished by scholars (Caudal 2005; Tenny 1994), while no distinction is usually made between argument-scale and inherent-scale homomorphism. Additionally, the constraints on selecting the types of argument enabling those three types of measuring remain unclear. One of the tasks of the present study is to verify the theoretical claim that some arguments have a role in expressing temporality.

Chapter 3

The Polish verb, Verbal Aspect and temporality

3.1 Theoretical approaches to PVA

In this chapter, I turn to the opposition between PFV and IPFV in Polish[1], which covers the whole verbal paradigm.

During the twentieth century, a certain development in the research on Slavic verbal aspect can be observed. The early studies of aspect were influenced by structuralism and the work of Jakobson (1971[1932]) which approached verbal aspect as a binary, privative opposition. One of the central question was markedness and the semantic invariant of the marked counterpart. Since the meaning of PFV seemed easier to define, it received the status of the marked counterpart – both morphologically and semantically. Holvoet (1989) distinguishes two directions in which this question develops in Polish linguistics. In the *temporal-deictic* direction (in Polish linguistics induced by the seminal work of Koschmieder 1934), the notion of verbal aspect is built upon temporal deixis and belongs to the same level of language as tense and mood. In the other direction (represented by Antinucci & Gebert 1957), the semantics of the predicate-argument structure is affected by PVA.

In time, the structuralist approach lost popularity. In other words, scholars stopped searching for invariants, and shifted towards describing the functions of PVA. The results of this approach are seen in the works of Śmiech (1971, 1986)

[1]The narrow group of biaspectual verbs, that is, verbs whose aspect cannot be evaluated, poses an exception and lies beyond the scope of this study. An extensive list of biaspectual verbs can be found in Perlin (2010).

or the relatively recent description of PVA in the descriptive grammar of Polish by Bartnicka et al. (2004). Since finding one invariant proved impossible, scholars concluded that PVA is not a unified mechanism, and its description requires accounting for several semantic features or levels.

Holvoet (1989) himself approaches PVA as a category with two levels: the temporal, tightly related to the temporal deixis, and the semantic, independent of temporal deixis, but referring to time as perceived by human beings. As to the temporal level, Holvoet (1989) follows Wierzbicka (1967) who claims that simultaneity is part of IPFV meaning, which PFV lacks. Duration in time and the existence of temporal bounds are to the temporal-semantic meaning of PVA. These two dimensions are realised in various ways for different types of predicate, so Holvoet does not define a single semantic invariant, but distinguishes a set of functions which are realised by particular aspectual values.

Similarily to other notions from the temporal domain, PVA can be also described in terms of scalarity. I focus on that matter in Section 3.7, having first characterised Polish verbs (Section 3.2), and described the morphology of PVA (Section 3.5). In the last part of the current chapter, I discuss particular levels of temporality in interaction with PVA.

3.2 Polish verbs

In the present work, I depart from the morphological template of the Polish verb where the verbal form consists of an obligatory stem and inflectional morphemes:

(PREFIX(ES)) + [ROOT+STEM SUFFIX]$_{stem}$ + INFLECTIONAL MORPHEMES + (REFL)

Finite forms of verbs which are studied in the present work are inflected for tense (see Section 3.8.2), person and number, and in the Past tense also for gender.

The notion of the verbal stem is described in Section 3.5.2, aspectual affixes are described in Section 3.5. The grammatical status of reflexive marker *się* is not entirely clear. *Się* oscillates between a clitic and a verbal affix (Fehrmann et al. 2010: 204), so it is not integrated into the verbal lexeme as much as, for example, in Russian. Futhermore, particular types of reflexive may block the canonical realisation of arguments in a sentence (Fehrmann et al. 2010: 204). The impact of valency reduction on PVA be relevant has not been subject to studies. Additionally, previous studies (Biskupska 2018; Kangasmaa-Minn 1984; Tommola 1986)

suggest that the correlates of PVA in Finnish might be related to the predicate-argument structure. Therefore, I start with an overview of this peculiar category, before discussing PVA in detail.

3.3 The Polish reflexive marker *się*

The reflexive marker *się* is usually considered polyfunctional. The reflexive typology of Fehrmann et al. (2010) in Slavic adjusted for Polish is given below:

Genuine reflexive – first and second argument are coreferential

(23) *Myję* *się.*
wash.IPFV.1SG REFL
'I wash myself.'

Reciprocal – the situation comprises two acts happening bidirectionally described by one verb. The agentive referent of the first act is coreferential with the non-agentive argument of the second act and vice versa. All referents are encoded as subject.

(24) *Całowali* *się* *na peronie.*
kiss.IPFV.PST.VIR3PL REFL on platform.LOC
'They kissed on the platform.'

Impersonal – the agentive argument is not realised (null subject), because the sentence represents a pattern (see Section 2.5.2) or the agent is hidden.

(25) *W Bawarii* *je* *się* *golonkę.*
in Bavaria.LOC eat.IPFV REFL ham.hock.ACC
'People in Bavaria eat ham hock.'

(26) *Na przyjęciu tańczyło* *się* *do* *rana.*
on party.LOC dance.acipfv.PST.NEU REFL until morning.GEN
'People danced until morning at the party.'

Reflexive passive – non-agentive argument is realised in the nominative subject while the agentive argument is expressed in the instrumental (27) phrase or as a prepositional phrase.

(27) *Szkoła buduje się rękami robotników.*
school build.IPFV REFL hand.PL.INS worker.PL.GEN
'School is built with hands of workers'

A particular type of reflexive passive is used to express generic meanings (28a). The scope of usages is narrower than the one of impersonal patterns as above (25) because the manner adverbial is obligatory.

(28) a. *Książka dobrze się czyta.*
book well REFL read.IPFV
'The book reads well.'

b. * *Książka czyta się w Bawarii.*
book read.IPFV REFL in Bavaria.LOC
('The book is read in Bavaria')

Involuntary State Constructions (ISC) are constructions with a dative experiencer and null subject.

(29) *Dobrze mi się czyta tę książkę.*
book me.DAT REFL read.IPFV this.ACC book.ACC
'I am reading this book with pleasure.'

(30) *Dobrze nam się tutaj mieszka.*
well us.DAT REFL here live.IPFV
'We feel we like living here.'

Antipassive – In comparison to the non-reflexive predicate (31) the second argument is either interpreted as arbitrary (32) or optionally realised as an oblique argument (33).

(31) *Powiem, jeśli pan podejmie się tej*
along.PFV.know.IPFV.1SG if sir under.PFV.take REFL this.GEN
*sprawy i podpiszemy **umowę**.*
matter.GEN and under.PFV.sign.~~IPFV~~.1PL agreement.ACC
'I will tell (that), if you, sir, undertake this matter and we sign the agreement. (Przepiórkowski et al. 2012)

(32) *Postawili u dołu datę,*
along.PFV.place.~~IPFV~~.PST.VIR.3PL at bottom.GEN date.ACC
podpisali się i
under.PFV.sign.~~IPFV~~.PST.VIR.3PL REFL and
wyszli, zamykając pokój.
out.PFV.go.~~IPFV~~.PST.VIR.3PL closing room.ACC
'They placed the date underneath, signed and left, closing the room.' (Przepiórkowski et al. 2012)

(33) *Podpisał się **na umowach**, kreśląc drżącą*
under.PFV.sign.~~IPFV~~.PST REFL on contract.LOC drawing shaking.INS
dłonią nieprawdopodobne kulfony.
hand.INS incredible.PL.ACC illegible.letter.PL.ACC
'He signed the contracts, by drawing incredibly illegible letters with his shaking hand.' (Przepiórkowski et al. 2012)

Decausatives the non-agentive argument is realised as subject and there is no causer

(34) *Czułem, że nasz czas **się** kurczy, ale w Projekcie*
feel.IPFV.PST.M.1SG that our time REFL shrink.IPFV but in project.LOC
Cheops zmieniało się niewiele.
Cheops from.~~PFV~~.change.IPFVPST.NEU REFL not.much
'I felt that our time was getting shorter, but there was not much change in the Cheops Project.' (Przepiórkowski et al. 2012)

Lexical reflexives – additionally to reflexives blocking argument realisation, a group of reflexives contributes only to the meaning of lexemes which must not

appear without the reflexive marker: *śmiać się* 'to laugh', *mądrzyć się* 'to try to be smart'.

3.4 Basic facts about PVA

The Imperfective applies prototypically to verbs denoting situations which last in time, and it does not specify the limits of lasting. It can appear in the following context:

(35) *Kup-owa-ł-e-m* *bułkę.*
Buy-IPFV-PST-M-1SG roll.ACC
'I was buying a/the roll.'

The Perfective moves the focus to attainment of some limit which implies the termination of the situation, while its lasting in time is secondary or irrelevant:

(36) *Kup-i-ł-e-m* *bułkę.*
Buy-PFV-PST-M-1SG roll.
'I bought a/the roll.'

The contrast between perfective and imperfective verbal lexemes (PERFECTIVES and IMPERFECTIVES) which contain the same root enables also finer distinctions. One of them is distinguishing between 'a quantum kick of the ball' (expressed with the perfective verb *kopnąć*) as in (37) and a situation called 'kicking a ball' (expressed with the Imperfective verb *kopać*) as in (38). In this case, the formal marking of this distinction is achieved with the morpheme *-nąś-* in the perfective verb.

(37) *Kop-nąś-ł-e-m* *piłkę.*
Kick-PFV.SEMEL-PST-M-1SG ball.ACC
'I kicked the ball.'

(38) *Kop-a-ł-e-m* *piłkę.*
Kick-IPFV-PST-M-1SG ball.ACC
'I was kicking a/the ball.'

Although the semantic difference between what is marked with Perfective and Imperfective is easiest to grasp by contrasting the usage of verbs sharing the same root, it would be wrong to say that all verbs in Polish form regular aspectual

pairs. However, the contrast in morphological structure is crucial for interpreting aspectual values and I describe it in the next section.

3.5 Morphological marking of PVA

3.5.1 The problem of morphological marking

Traditionally, two types of marking PVA are distinguished: *prefixisation* and *suffixisation*. Prefixes appear before the root of the verb, suffixes appear after the root. Concluding from the template (Section 3.2), PVA does not need to have surface marking.[2] The problem of markedness cannot be solved easily, because the morpheme and categories of PVA do not correspond directly. As I explain below, both prefixes and suffixes are polyfunctional and they do not form any clearly unified system.

3.5.2 Verbal stem

The paradigm of Polish verbal lexemes must always account for two variants of the stem.[3] The infinitive, Past tense, passive participle, and past participle usually share one stem variant, while (part of) the Non-past tense, imperative, active participle and present participle paradigm usually share another stem variant, as shown below:

(39) a. *koch-a-ć*
love-SUFF-INF

b. *koch-aj-ą*
love-SUFF-3PL

(40) a. *sol-i-ć*
salt-SUFF-INF

b. *sol-ø-ą*
salt-ø-3PL

[2]Nevertheless, some scholars argue that in verbs like *rzuc-i-ć* 'to throw (a discrete situation)' the morpheme *-i-* is the marker of the Perfective, while the morpheme *-a-* in the Imperfective *rzuc-a-ć* 'to throw, a continuous act or a series of discrete or continuous acts' is an aspectual marker (Wróbel 1998: 566).

[3]It is probably not possible to describe modern Polish verbal stems both simply and consistently due to suppletion and alternation. Consequently, no single, well-acknowledged solution is given. The account given here is very simplified, because verbal morphology is not in the focus.

(41) a. *trze-ø-ć*
rub-ø-INF

b. *tr-ø-ą*
rub-ø-3PL

The stem consists of the root and the stem suffix. The stem suffix may be realised only in one variant of the stem, or not realised in the paradigm. I now discuss particular types of suffix in relation to the aspectual properties of verbal lexemes. In the description, I use the infinitive form, and whenever relevant, I provide both suffixes.

3.5.3 Non-prefixed perfectives

Simplex perfectives

The group of perfective, non-prefixed verbs in Polish forms a very short list (Łazorczyk 2010: 16) : *chwycić* 'to grab', *chybić* 'to miss a goal', *czepić* 'to cling to, to stick to', *dać* 'to give', *kupić* 'to buy, to purchase', *lec* 'to lie down', *paść* 'to fall down', *puścić* 'to let loose', *rzec* 'to say, announce', *ruszyć* 'to start', *rzucić* 'to throw', *skoczyć* 'to jump', *stawić* 'to place', *strzelić* 'to shoot, to fire a gun', *trafić* 'to reach, to hit, to find one's way'. I call these verbs SIMPLEX PERFECTIVES. Most verbs in this group have at least one variant of stem ending with the suffix *-i-* or *-y-* or do not realise the stem suffix.

Semelfactives

The separate semelfactive suffix *-nąˊ/-n-* (which should not be confused with the homographic translative suffix described in Section 3.5.4) is an explicit Perfective marker and it has an additional meaning, "the smallest unit of action, a quantum of action" Holvoet (1989: 51), as presented in (37) (e.g. *walnąć* 'to thud', miauknąć 'to meow', *mrugnąć* 'to wink').

3.5.4 Non-prefixed imperfectives

Simplex imperfectives

Non-deverbal verbs described below I call SIMPLEX IMPERFECTIVES. The infinitive stem variant of imperfective non-prefixed verbs consists of a vowel: *lub-i-ć* 'to

like', *ż-y-ć* 'to live', *pis-a-ć* 'to write', *wiedzi-e-ć* 'to know', or in the case of verbs describing directed motion with suffix *-nąˊ-n- pły-nąˊ-ć* 'to swim somewhere'.

A small group of simplex verbs are suppletive verbs like *jeść* 'to eat', *iść* 'to go', *kraść* 'to steal' which do not have a stem suffix.

Imperfective stem suffixes also include *-owa-/-uj-* (*głos-ow-a-ć* 'to vote'), *-izowa-/-izuj-* (*kolonizować*) 'to colonise', and *-yzowa-/-yzuj-* (*terroryzować* 'to terrorise').

Translatives

Deadjectival imperfective verbs can be formed with the suffix *-nąˊ-/-n-* (*blednąć* 'to become more pale'), or with *-ie-/iej-* (*widnieć* 'to be visible somewhere or to become brighter', *ładnieć* 'to become beautiful or more beautiful'). These verbs share a meaning 'to be or become as described by the adjective from which the verb is derived'. I call them TRANSLATIVES.[4]

Deverbal non-prefixed imperfectives

Within deverbal, nonprefixed imperfectives three groups can be distinguished. HABITUALS are derived from other imperfectives and express habits: *pi-j-a-ć* 'to have a habit of drinking' in comparison to *pi-ć*, *ch-**a**-dz-**a**-ć* 'to have a habit of going somewhere' in comparison to *ch-**o**-dz-**i**-ć* 'to walk'; *siad-ywa-ć* 'to have a habit of sitting down somewhere' in comparison to *siadać*, *pis-ywa-ć* 'to have a habit of writing'.

Morphologically similar changes apply to verbs of undirected movement. also derived when contrasted with the imperfective verbs of directed movement: *pły-wa-ć* in comparison to *pły-nąˊ-ć*, *l-a-t-a-ć* in comparison to *leci-e-ć*. In the third group, all simplex perfectives, but *rzec* have a non-prefixed imperfective correlate: *chwytać, chybiać, czepiać, dawać, kupować, padać, puszczać, ruszać, rzucać, skakać, stawiać, strzelać, trafiać*. I will call verbs belonging to the latter two groups BARE imperfectives.

[4] I apply this term after Uralic tradition, see Section 4.2.1

3.5.5 Prefixed verbs

Inventory of prefixes in Polish

In Polish the prefixes: *de(z)-*, *kontr-*, *niedo-*, *re-* *współ*, and *przeciw-* are not considered aspectual markers (Wróbel 1998: 564; Łazorczyk 2010: 24; Biskupska 2018: 55).

Depending on the treatment of some prefixes, scholars distinguish between 26 (Bartnicka et al. 2004), 17 (Biskupska 2018; Śmiech 1986) and 18 (Łazorczyk 2010)[5] aspectual prefixes which carry a meaning of some spatial relation. Most of them correspond synchronically with the form and meaning of some preposition (rows 1–14 in Table 3.1). The exhaustive list is given in Table 3.1 (based on Łazorczyk 2010: 22–23 and Łaziński 2011: 234). Prefixed verbs are perfective (as long as they do not receive an additional imperfective suffix) and they are derived from simplex perfectives and simplex imperfectives. Since the population of simplex imperfectives is bigger than the population of simplex perfectives, prefixed verbs are more frequently derived from simplex imperfective stems. In some cases, the non-prefixed base does not exist and the prefixed verb is derived directly from an adjective or noun as in *przy-bliż-y-ć* 'to bring closer, to magnify' synchronically related to the adjective *bliski* 'near, close'.

prefix	meaning	non-prefixed	prefixed derivate
do-	'to, into, towards, till, until´	*dać* 'give'	*dodać* 'to add'
na-	'on, onto'	*chlapać* 'to splash'	*nachlapać* 'to splash on something'
nad(e)-	'above, over'	*pisać* 'to write'	*nadpisać* 'to overwrite'
o-	'around, about'	*snuć* 'to plot'	*osnuć* 'to wrap with a thread'
od(e)-	'away, from'	*spać* 'to sleep'	*odespać* 'to sleep off'
po-	'over, after'	*bielić* 'to whiten'	*pobielić* 'to whitewash'
pod-	'under, below, beneath'	*pisać* 'to write'	*podpisać* 'to sign'

[5]Łazorczyk (2010) treats the prefixes*ob-* and *o* separately, whereas most scholars treat them as variants of the same prefix (cf. Biskupska 2018: 55). In the present work they are considered separately, making it possible to trace the differences in quantitative patterns.

przed(e)-	'in front of, before, prior to'	*stawić* 'to place'	*przedstawić* 'to introduce'
przy-	'next to, at'	*kuć* 'to forge'	*przykuć* 'to chain, to nail'
u-	'at, by'	*dramatyzować* 'to panic'	*udramatyzować* 'to adjust for a stage'
w(e)-	'in, inside, into'	*czytać* 'to read'	*wczytać* 'to read in'
z(e)-/ s- / ś-	'from, off, with'	*skoczyć* 'to jump'	*zeskoczyć* 'to jump down'
za-	'behind, instead, next to'	*malować* 'to paint'	*zamalować* 'to paint over'
ob(e)-	'around, about'	*smarować* 'to slush'	*obsmarować* 'to dish the dirt on sb.'
prze-	'through, throughout, across, by'	*czytać* 'to read'	*przeczytać* 'to finnish reading'
roz(e)-	'from the centre in many directions, dispersion'	*śmiać się* 'to laugh'	*roześmiać się* 'to burst laughing'
wy-	'out, outwards'	*rzucić* 'to throw'	*wyrzucić* 'to throw away'
wz(e)- / wez- / ws- / wes-	'upward'	*tchnąć* 'to breath into'	*westchnąć* 'to sigh'

Table 3.1: Polish perfective prefixes

Prefixes and productivity

Most prefixes are productive and they combine with new stems. Łazorczyk (2010: 27 – 28) attests the examples for all prefixes but *nad-*, *przed-* and *wz-* in combination with stems related to information technology *klikać* 'to klick' and *blogować* 'to blog'. On the other hand, usually only certain combinations of prefixes are possible with the given root, and it is challenging to form any general rules about them.

In many cases the morphological structure of prefixed verbs is visible, although the etymon is not used any more. Usually several prefixed cognates can be compared, for example, to those derived from Proto-slavic **ględati* 'to watch': *o-glądać* 'to watch, to inspect, to behold', *wy-glądać* 'to look, to appear', *prze-*

glądać ‘to glance through’.

Secondary imperfectives

Deriving new imperfective verbs from prefixed perfectives is possible too. In the literature, such verbs are called SECONDARY IMPERFECTIVES (cf. Dickey 2000; Łazorczyk 2010). The morphological change in the stem is similar to deverbal imperfectives. The stem suffix of the derivative base is replaced with the new, more complex one (Wróbel 1998: 565), for example, *przepis-ywa-ć* in comparison to *prze-pis-a-ć* or *przepi-ja-ć* in comparison to *prze-p-i-ć*. The suffix *-owywa-/-owuj-*[6] is also used when the suffix stem of the prefixed perfective verb is *-owa-/-uj-* (*przebud-ow-ywa-ć* in comparison *prze-bud-owa-ć* ‘to rebuild’).

Prefix stacking

Prefixes can be stacked in Polish, attaching in front of other prefixes. This largely occurs in secondary imperfectives, as in the verb *po-od-po-wiadać* ‘to DISTRIBUTIVE.answer’, although, as noted in Łaziński (2011: 234), in most cases the original stem becomes blurred. In the example above only the first two prefixes (*po-* and *do-*) are synchronically lexically motivated. Nowadays, only prefixes *po-* and *na-* have an unrestricted distribution in the context of prefix stacking (see Section 3.6.2).

3.5.6 Formal status of PVA

The irregular character of PVA markers causes disagreement among scholars about the formal status of aspectual markers in Polish, or more broadly speaking in Slavic languages. Grzegorczykowa et al. (1984) suggested that aspectual suffixes belong to verbal inflection, while prefixes belong to verbal derivation. This approach implies that some verbs are grammatical variants of one lexeme, but other verbs which share the root represent different lexemes. A similar approach is followed in Bartnicka Bartnicka et al. (2004: 328–9).

Laskowski (1998b) offers a more unitary solution is offered – each verb is treated as a unique lexeme and PVA is a classificatory category analogical to noun

[6]The derivation process could be considered in terms of deriving a new stem from the base stem, but this would require introducing an additional notion of stem vowel (e.g. in Łazorczyk 2010). The main disadvantage of such an approach is the increased complexity of the description.

gender. The approach has recently been defended by Wiemer & Seržant (2017), and I follow it here.

3.6 Verbal affixes and lexical verbal meaning

3.6.1 Aktionsart

As observed above, apart from aspectual contrast between IPFV and PFV in derivation from simplex verbs, affixes carry a semantic surplus which contributes to the lexical verbal meaning. Some prefixes only modify the meaning of the verb according to the spatial meaning of the cognate preposition. In these cases, prefixed verbs often occur in the clause with the corresponding preposition of the argument:

(42) a. *Heinz* ***o****powiedział* ***o*** *dwóch ciałach*
Heinz about.PFV.tell.~~IPFV~~.PST about two.PL.LOC body.PL.LOC
znalezionych nad Wisłą.
found.PL.LOC by Vistula
'Heinz told about two bodies found at the Vistula river' (Przepiórkowski et al. 2012)

b. *Edyta* ***o****powiedziała wszystko ze szczegółami.*
Edyta about.PFV.tell.~~IPFV~~.PST everything.ACC with detail.PL.INS
'Edyta reported everything in detail.' (Przepiórkowski et al. 2012)

On the other hand, prefixes often develop some additional, functional meanings besdies the spatial or prepositional ones. In aspectual studies, this phenomenon is called AKTIONSART. The term was first used for by Sigurd Agrell Agrell (1908) to describe the semantics of the rich inventory of Polish aspectual morphemes.[7]

[7] Aktionsart is often confused with *lexical aspect* or *types of situation*, which is a direct translation of the German word. While the original term refers exclusively to the meaning of verbal affixes in Slavic languages, the latter two terms apply either to the inherent actional properties of the whole verb in languages which do not encode a distinction similar to PVA in the whole verbal paradigm or to the actional properties of whole predicates. Many actional properties can be derived from Aktionsart which is probably the source of confusion. The classic general works describing lexical aspect and types of situation are: Vendler (1957), Comrie (1976), Mourelatos (1978), for Polish Laskowski (1996, 1998a), for Finnish Leino (1991) and Hakulinen et al. (2004).

Contemporary descriptions of prefix meanings and Aktionsart in Polish are presented in Bartnicka et al. (2004: 330-339) and Wróbel (1998: 545-568). The former adopts a form-based model – each morpheme is described separately with all its possible meanings. In the latter, the functional meaning is a distinct factor to which all possible morphemes are assigned. Śmiech (1986) only considers prefixation and takes a form-based approach. His detailed, atomistic descriptions do not provide any general, unified conclusions which help explain the link between PVA and Aktionsart.

In her relatively recent dissertation, Łazorczyk (2010) elaborates on some issues of morphology of PVA within the frame of generative grammar and with some reference to diachronic studies, but it does not provide a full overview of Aktionsart. The main focus is placed on the prefixes and the secondary imperfective prefixes (imperfective prefixes in the prefixed, imperfective verbs, and imperfective morphemes built around the consonant *w*). The author considers prefixes and suffixes in question to be the markers of telicity (prefixes) or detelicisers (secondary suffixes).[8]

Aktionsart applies to the very frequent prefixes, such as *po-*, *na-*, *za-*, and *prze-*. I now compare their spatial meanings and the types of Aktionsart they are capable of expressing.

3.6.2 The prefix *po-*

The prefix *po-* is the least lexicalised prefix. It keeps the original spatial meaning only in combination with a verbal root which encodes the meaning related to motion along a surface. This is visible in the examples: *pokryć* 'to cover with something', *pobielić* 'to cover with a white substance', *pobiec* 'to run along some path'.

A different meaning of *po-* refers to a 'portion of time', *poczytać* 'to read a bit, for a while', *potańczyć* 'to dance a bit, for a while'. The usage of the prefix *po-* in this function is called DELIMITATIVE (DELIM).

Po- might have inchoative (INCH) meaning; an independent situation is a phase of another situation represented by the verbal root. This function of *po-* applies mainly to situations describing the emotions like *pokochać* 'to start loving' or *polubić* 'to start liking'.

In many contexts, in particular in prefix stacking, *po-* marks either a distributional motion (one by one, one to many), for example, *po-roz-dawać* 'to dis-

[8] In this case, following Borer (2005), telicity is understood as satisfying the condition of not being cumulative or not being divisible by the predicate in an argument structure.

tribute', *po-z-bierać* 'to collect', *po-za-stanawiać* 'to think, to consider one by one, in portions of time'. Such a usage of the prefix is called DISTRIBUTIVE (DISTR).

3.6.3 The prefix *na-*

The prefix *na-* often keeps its lexical meaning[9] as in the verb *napisać* 'to write down', but in many cases its meaning can be better understood in terms of accumulation,as in *nałowić ryb* 'to catch plenty of fish' or *nagadać się* 'to talk a lot, to satisfy the need for chatting'. In order to distinguish between the motion *onto* and the multiple motion, the second usage is called ACCUMULATIVE (ACCUM). Accumulative *na-* also has unrestricted distribution in prefix stacks.

3.6.4 The prefix *za-*

The spatial semantic contribution of *za-* is *behind or behind the border*, and it is realised in the verbs *zakryć* 'to cover', *zalać* 'to pour on', *zatkać 'to clog up, to plug up)* where the subject covers the object. A similar relation is visible in the verbs *zajechać* 'to drive by', *zasiąść* 'to sit down to' a similar relation is visible.

The problems pose verbs such as: *zakochać się* 'to fall in love', *zaśpiewać* 'to perform a song', *zawrzeć* 'to cook to a full boil'. Śmiech (1986) suggests that those verbs fuse the meaning of the verb *zaczać* 'to start, to begin' and of the root.

Alternatively, Bartnicka & Satkiewicz (2000: 273) and Tabakowska (2003: 171) consider the prefix *za-* as a *pure aspectual marker* that does not contribute any surplus to the lexical meaning of the verb.

However, in many cases, the presence of *za-* may be explained with the fact that the subject becomes covered either physically, as in *zatonąć* 'to sink down', or metaphorically *zadumać się* 'to sink in thoughts', *zakochać się* 'to fall in love'. Similar points were made by Dąbrowska (1996) and Bacz (2005).

In a sense, all presented points may be true, for certain groups of verbs, according to the type of arguments they require. Verbs keeping the spatial meaning require either an object which becomes covered, or a prepositional directional argument: *zajechać przed* + ACC 'to drive by' , *zasiąść do* + GEN 'to sit down to'.

Verbs where the subject becomes covered need an argument with a preposition *w* 'into', or occasionally they occur with the reflexive marker *się*. Although the

[9]According to the corpus-based study of Łaziński (2011), among all prefixes *na-* occurs most frequently in clauses with the corresponding preposition.

marker is lexical (verbs without *się* are not transitive) they indicate that the patient or mover is realised as subject.

Verbs like *zatańczyć* 'to dance', *zaśpiewać*, *zagrać* 'to play' cannot be explained as lexically motivated. Object argument, if it appears, is partially semantically redundant. It is obvious that if one dances, one dances a dance, and if one plays, one plays music or a game. Thus, the object of those verbs is in fact an inherent-scale argument (see Section 2.6.4).

In that case *za-* introduces a limit, similarly to DELIM.*po-*. The difference between the two prefixes is that while *po-* seems to prefer delimiting straightforwardly over the duration (unless an interval is given, the implicit interval is *a while*), *za-* is used in a delimitative meaning when the inherent-scale argument is possible.

A similar function of *za-* can explain the meaning of verbs *zadzwonić* 'to call, to ring', *zamówić* 'to order', and *zapłakać* 'to cry'. Admittedly, those verbs are intransitive, but interestingly, all are derived from verbs meaning 'to produce sound', thus analogues to *zaśpiewać* 'to sing'. This meaning could be explained as a 'unit of action' very similar to the function of a semelfactive morpheme.

3.6.5 The prefix *prze-*

In accordance with the lexical meaning of *prze-*, prefixed verbs an require object argument to which the real or metaphorical relation *through* applies. This may refer to movement, for example, *przekroczyć ulicę* 'to cross the street'. Metaphorically, it functions like the English *go through*, for example, *przeliczyć dzieci* 'to count children', *przeanalizować dane* 'to analyse data'.

The PERDURATIVE (PERDUR) function of *prze-* is realised when object argument refers to an interval, and thus has a durational reference. The object can explicitly name the interval, for example, *przesiedzieć na zebraniu* ***godzinę*** 'to sit for an hour in the meeting', *przetańczyć* ***całą noc*** 'to dance all night long', or implicitly, for example, *przespać* ***cały wykład*** 'to sleep through the whole lecture'. *prze-* is thus a DISCRETISER, similarly to DELIM.*po-* and *za-* in the function of delimitative or semelfactive.

3.7 Defining PVA in terms of scalarity

Following the scalar approach to temporality presented in Section 2.6, I discuss now the opposition between PFV and IPFV in terms of scale. PVA determines the

type of scale on which the TR to which TSIT is assigned is measured. PFV enables assignment to units measured on discrete scales, so the temporal unit cannot be quantised, while IPFV deals with temporal units measured on continuous scales.

(43) a. *(Wczoraj) czytałam książkę*
Yesterday read.IPFV.PST.F.1SG book.ACC
'(Yesterday) I was reading/read a/the book.'

b. *(Wczoraj) prze-czytałam książkę*
Yesterday through.PFV-read.~~IPFV~~.PST.F.1SG book.ACC
'(Yesterday) I read through the book.'

In (43a) the TR localised in *yesterday* should be measured on the continuous scale. It may be quantised so the situation may be assigned to all points from this temporal unit but it may be assigned to one point as well. This is why the verb in this sentence cannot be unambiguously interpreted without broader context and easily translated into English as Progressive.

In (43b) the temporal unit is measured on a discrete scale, so it can take only a particular value within the temporal unit *yesterday*.[10]

The difference is visible when different types of perfectives and imperfectives are tested with an adverbial phrase marking the beginning and the end of a temporal interval:

(44) a. ***Od stycznia do kwietnia*** *chadzałam na*
From January.GEN until April.GEN walk.IPFV.HAB.PST.F.1SG on
jogę.
yoga.ACC
'I used to go to yoga classes from January until April.'

b. * ***Od stycznia do kwietnia*** *kopnęłam*
From January.GEN until April.GEN kick.PFV.SEMEL.PST.F.1SG
piłkę.
ball.ACC

[10]The predicate *przeczytałem* in (43b) could be replaced with other prefixed verbs which do not differ as to PVA but in meaning. In particular the predicate *poczytałem* with the delimitative prefix *po-* could be used. How to treat delimitative Aktionsart in the scalar approach is discussed in Section 3.8.3.

(‘I kicked once a/the ball from January until April.’)

c. ***Od stycznia do kwietnia*** *mieszkałam w*
From January.GEN until April.GEN live.IPFV.PST.F.1SG in
Warszawie.
Warsaw.LOC
‘I lived in Warsaw from January until April.’

d. * ***Od stycznia do kwietnia***
From January.GEN until April.GEN
prze-czytałam książkę.
through.PFV-read.~~PFV~~.PST.F.1SG book.ACC
(‘I read through a book from January until April.’)

e. ***Od stycznia do kwietnia*** *prze-kopywałam*
From January.GEN until April.GEN through.~~PFV~~-dig.IPFV.PST.F.1SG
ogródek.
garden.ACC
‘I dug through the garden from January until April.’

f. * ***Od stycznia do kwietnia***
From January.GEN until April.GEN
po-kopałam w ogródku.
DELIM.PFV-dig.~~IPFV~~.PST.F.1SG in garden.LOC
(‘I dug in the garden a bit from January until April.’)

The temporal unit *from January until April* is an interval so it has a continuous structure. The Perfective verbs are used incorrectly because they cannot be assigned to all the points of this temporal unit.

The sentences containing perfectives could be modified as follows:

(45) a. *Od stycznia do kwietnia* ***raz*** *kopnęłam*
From January.GEN until April.GEN once kick.PFV.SEMEL.PST.F.1SG
piłkę.
ball.ACC
‘I kicked a/the ball once from January until April.’

b. *Od stycznia do kwietnia* ***raz***
From January.GEN until April.GEN once
przeczytałam książkę.
through.PFV.read.~~IPFV~~.PST.F.1SG book.ACC
'I read a/the book once from January until April.'

c. *Od stycznia do kwietnia pokopałam*
From January.GEN until April.GEN DELIM.PFV.dig.~~IPFV~~.PST.F.1SG
raz *w ogródku.*
once in garden
'I dug once a bit in the garden once from January until April.'

Now in all these sentences PFV is justified because it is said that the continuous temporal unit 'from January to April' contains one discrete temporal unit. *Raz* 'once' could be replaced with some higher number of instance, as long as it would imply a finite number. This is why the adverbial phrases equivalent to *twice, many times* would be compatible, but *unceasingly* would not.

In (45b) quantification of the direct object (one book), would give the same effect as specifying the frequency.

3.8 Interaction between PVA and other markers of temporality

3.8.1 Basic types of interaction

In this section I discuss how PVA contributes as a marker in the subdomains of temporality distinguished in Chapter 2. Some clear differences between IPFV and PFV within temporal localising, durative temporalisation and pluractionality are visible in the examples below:

Only imperfectives may assign TSIT to the current present:

(46) a. *Właśnie rzucam piłkę.*
right.now throw.IPFV.1SG ball.ACC
'I am throwing the ball right now.'

b. * *Właśnie rzucę piłkę.*
Right.now throw.PFV.1SG ball.ACC
('I will throw the ball right now.')

Genuine durative expressions may appear only with imperfectives, while perfectives may take only frame adverbials (see Section 3.8.3):

(47) a. *Rozwiązywałam zadanie piętnaście minut.*
dispersion.~~PFV~~.bind.IPFV.PST.F.1SG task.ACC fifteen minute.PL.GEN
'I solved the task for fifteen minutes.'

b. *Rozwiązałam zadanie* ***w*** *piętnaście minut.*
dispersion.PFV.bind.~~IPFV~~.PST.F.1SG task.ACC fifteen minute.PL.GEN
'I solved the task in fifteen minutes.'

Only imperfective verbs may express an unspecified number of repetitions:

(48) a. *Ciągle kupowałam tutaj kawę.*
unceasingly buy.IPFV.PST.F.1SG here coffee.ACC.
'I bought my coffee here all the time.'

b. * *Ciągle kupiłam tutaj kawę.*
unceasingly buy.PFV.PST.F.1SG here coffee.ACC.
('I bought my coffee here all the time.')

In the following, I elaborate on such interactions with regard to different temporal subdomains.

3.8.2 PVA and temporal localising in Polish

Morphological tense marking

When describing the Polish tense system, the Perfective and the Imperfective paradigms must be kept apart.

Perfectives appear only in the Past tense marked with the *l-participle* (LPTCP) and in the morphologically unmarked the Non-past tense. Imperfectives form the Past and the Non-past tense in the same way as perfectives, but additionally they appear in the future analytical form, which consists of the auxiliary formed from the verb *być* 'to be' and the infinitival imperfective complement or the LPTCP. The last variation does not seem to arise from the semantics, and scholars prefer to explain that difference diachronically with the different ages of the constructions in question (Błaszczak et al. 2014 after Whaley 2000).

Absolute tenses

The tense system of perfectives is bipartite, and for imperfectives tripartite. This leads to differences in the scope of temporal reference, which mainly relate to the usage of the Non-past and to future temporal reference.

The Non-past form of imperfectives may refer to the current present, as in (46a). In other words, this form may be used to refer to a TSIT assigned to a TR identical with a TU. This is the main difference in comparison to perfectives.

In the case of perfectives, the Non-past tense must refer to a TSIT associated with a TR posterior to a TU. It can be either a moment immediately following a TU as in (49a) or some more distant point in the future as in (49b).

(49) a. *Wyjdę* ***teraz*** *do sklepu.*
out.PFV.go.1SG now to shop.GEN
'I am leaving to go to the shop in a moment'

b. *Wyjdę* *do sklepu* ***za*** ***dwie*** ***godziny***.
out.PFV.go.1SG to shop.GEN behind two.ACC hour.PL.GEN
'I will leave to go to the shop in two hours'

The imperfective analytical forms are used to anchor TSIT to TR posterior to TU.

(50) *...zapraszamy* *na jutrzejszy* *spektakl.*
behind.PFV.request.~~IPFV~~.1PL on tomorrow.ACC show.ACC
Występować *będzie* *słynna* *Beba Mazeppo.*
out.~~PFV~~.walk.IPFV.INF AUX.3SG famous Beba Mazeppo
'Welcome to the performance tomorrow. The famous Bebba Mazeppo will be performing.' (S1306)

The Non-past Imperfective may also have the future reference. Little research has been done on the differences between the usage of tense-aspect forms and the type of future temporal reference. Holvoet (1989) and Kochańska (2002, 2007) explain the differences with the speaker's level of certainty about whether situations will occur in future. According to Kochańska (2002: 371) situations which are only virtual (not scheduled, planned or regularly occurring) are aspectually more restricted:

(51) a. *Jutro skończą /kończą się wakacje.*
tomorrow PFV.end.~~IPFV~~.3PL /end.IPFV.3.PL REFL vacation
'Tomorrow our vacation will end/end.' (Kochańska 2002: 370)

b. *Jutro skończy /?kończy się nam herbata.*
tomorrow PFV.end.~~IPFV~~ /end.IPFV REFL we.DAT tea
'Tomorrow we will run out of tea.' (Kochańska 2002: 370)

While the end of a holiday is unavoidable, the statement about tea is only a forecast. Additionally, it is also beyond the speaker's control (either mental or physical). I return to this problem when discussing temporal quantification in Polish.

Taxis

Absolute tense marking is relatively simple in Polish. Taxis is not grammaticalised in absolute-relative tenses similar to the Pluperfect. Therefore, the five possible combinations of tense forms and PVA play an important part in taxis distinctions.

Holvoet (1989) who does not analyse taxis as a whole, but only the modifications caused by aspect in the tensed sentences he calls *referentiality* (pol. *referencjalność*) claims that one function of IPFV is marking simultaneity. Śmiech (1971) draws a similar conclusion, based on a very detailed analysis of possible taxis-like combinations between two clauses. Accordingly, in the temporal subordinate clauses, IPFV is the marker of simultaneity (52a), while PFV marks sequentiality (52b):

(52) a. *Kiedy matka przygotowywała kolację, dzieci*
when mother at.~~PFV~~.cook.IPFV.PST.F.3SG dinner.ACC child.PL
oglądały bajkę.
around.~~PFV~~.watch.PST.NVIR.3PL fairy.tale.ACC

'While mother prepared the dinner, the children were watching a fairy tale.' (Łuczków 2013: 4)

b. *Kiedy mnie zobaczył,*
when me.ACC from.PFV.around.~~PFV~~.see.~~IPFV~~
postąpił pół kroku w moją stronę
along.PFV.step.~~PFV~~.PST half step.GEN in my.ACC side.ACC
'Having seen me, he took half a step towards me.'(S887/888)

The usage of lexical items such as conjunctions, adverbs, or anaphoric expressions plays an important role in relative ordering of TSITs. Although the natural order of two clauses in PFV with a coreferential subject usally forms a sequence (53a), two clauses in IPFV are interpreted as simultaneous (53b):

(53) a. *Giugiu pokiwał rozumiejąco głową i*
Giugiu DELIM.PFV.nod.~~IPFV~~.PST understandingly head.INS and
zapalił papierosa.
behind.PFV.smoke.~~IPFV~~.PST cigarette.ACC
'Giugiu nodded his head as a sign of understanding and lit a cigarette.' (S1287/1288)

b. *Giugiu kiwał rozumiejąco głową i palił*
Giugiu nod.IPFV.PST understandingly head.INS and smoke.IPFV.PST
papierosa.
cigarette.ACC
'Giugiu nodded his head as a sign of understanding and smoked a cigarette.'

Most clauses contain additionally some expressions as *potem* 'then, afterward', *wtedy* 'at that moment', *najpierw* 'first', *później* 'later'. This is because two clauses with no coreferential subject will usually be interpreted as parallel or overlapping regardless aspectual values of PVA as in (54), while the acceptability of clauses with distinct aspectual values but coreferential subjects as in (55) is questionable:

(54) Giugiu **(po)kiwał** rozumiejąco głową, a Wolfgang **(za)palił** papierosa.

'Giugiu noded with his head as a sign of understanding and Wolfgang lit /smoked a cigarette.'

(55) ? Giugiu **pokiwał** rozumiejąco głową i **palił** papierosa.

Laskowski (1998b: 174) lists three types of taxis: simultaneity, anteriority and posteriority. These do not have separate markers, but are expressed differently in different contexts for example by tense, aspect, or lexically (conjunctions, temporal particles, or adverbs). Tense in the function of taxis appears primarily in subordinate object clauses, as shown below:

(56) a. *Anna wiedziała, że Adam **czytał** jej list.*
Anna know.IPFV.PST.F.3SG that Adam read.IPFV.PST her letter.ACC

b. *Anna wiedziała, że Adam **przeczytał***
Anna know.IPFV.PST.F.3SG that Adam through.PFV.read.~~IPFV~~.PST
jej list.
her letter.ACC

'Anna knew that Adam had read her letter.'

(57) *Anna wiedziała, że Adam **czyta** jej list.*
Anna know.IPFV.PST.F.3SG that Adam read.IPFV her letter.ACC
'Anna knew that Adam was reading her letter.'

(58) a. *Anna wiedziała, że Adam będzie czytał jej*
Anna know.IPFVPST.F.3SG that Adam AUX read.IPFV.LPTCP her
list po obiedzie.
letter.ACC after lunch.LOC

b. *Anna wiedziała, że Adam przeczyta jej*
Anna know.IPFVPST.F.3SG that Adam through.PFV.read.~~IPFV~~ her
list po obiedzie.
letter.ACC after lunch.LOC

'Anna knew that Adam will read/will be reading her letter after lunch.'

In (56) the Past tense in the subordinate clause refers to some unit of time anterior to the TSIT of the main clause, analogically, in (57) the Non-past tense

(and therefore IPFV), in the subordinate clause refers simultaneity with the TSIT of the main clause, and in (58) the usage of analytical future or PFV allows referring to the time posterior to the TR of TSIT of the main clause.

The verbal forms specialised in marking taxis are in Polish participles.[11]

Lexical temporal localising expressions

Lexical temporal localising completes the localising achieved by the tensed form of the verb. Expressions such as *today, nine o'clock*, names of events like *breakfast*, and names of festivals like *Christmas* help to localise TSIT on the time axis. The TR to which the TSIT is assigned can be pointed independently (59) or at a distance from some other point (60). It may be identical with the TU (61).

(59) *Rozpoczynamy* *obrady* ***o dziewiątej***.
dispersion.~~PFV~~.begin.IPFV.1PL proceeding.PL.ACC at nine.LOC
'We are beginning the proceedings at nine o'clock.'

(60) *Rozpoczynamy* *obrady* ***za*** ***godzinę***.
dispersion.~~PFV~~.begin.IPFV.1PL proceeding.PL.ACC behind hour.ACC
'We are beginning the proceedings in an hour.'

(61) *Rozpoczynamy* ***teraz*** *obrady.*
dispersion.~~PFV~~.begin.IPFV.1PL now proceeding.PL.ACC
'We are beginning the proceedings now.'

In Polish, temporal expressions participating in temporal localising have three forms:

- bare adverb (e.g. *jutro* 'tomorrow')
- noun phrase (e.g. prepositional noun phrase *w styczniu* 'in January', Genitive noun phrase *pierwszego stycznia* 'on January the first', Instrumental noun phrase *wiosną* oun phrase *pierwszego stycznia* 'in the spring')
- temporal particle (e.g. *już* 'already').

[11] The Polish participle system is nevertheless determined by PVA. The active adjectival participle and contemporary adverbial participle are formed from the imperfectives, anterior adverbial participle is formed only from the perfectives.

In the examples presented above, the TR of the TSIT was parallel to the named temporal unit, but the TR may also be anterior (62) or posterior (63). Both aspects may occur together with temporal adverbials.

(62) ***Przed obiadem*** *szłam /poszłam na*
before lunch.INS go.IPFV.PST.F.1SG /along.PFV.go.~~IPFV~~.PST.F.1SG onto
pocztę
post.ACC
'I was going /went to the post office before lunch.'

(63) ***Po wojnie*** *Warszawę odbudowywano*
after war.LOC Warsaw.ACC from.~~PFV~~.build.IPFV.~~IPFV~~.IMPS
/odbudowano.
/from.PFV.build.~~IPFV~~.IMPS
'Warsaw was being rebuilt /was rebuilt after the war.'

In the case of PFV, the TSIT is assigned to a discrete unit(s) within the TR, while IPFV allows for assignment to from one to all units of the TR, as shown in (43).

3.8.3 Durative temporalisation

Lexical expressions

Durative temporalisation is achieved by specifying the length-related properties of the TR to which the TSIT is assigned in the sentence. It is realised in Polish by adverbials:

- prepositional noun phrases *przez* + ACC (e.g. *przez godzinę* 'through the night')
- accusative noun phrase (e.g. *cały tydzień* 'all week')
- specifying qualitatively the lasting with an adverb (e.g. *długo* 'long', *krótko* 'short')

Some temporal expressions combine both durative temporalisation and temporal localising, for example when specifying the boundary conditions in adverbial noun phrases (*od wczoraj do jutra* 'from yesterday till tomorrow').

Concurrence restrictions between durative adverbials and PVA

The normal perception of an interval as a unit is continuous, therefore mainly situations referred to in imperfectives allow durative temporalisation.

Perfectives are assignable to discrete units, so normally they are not an object of durative temporalisation, but they occur concurrently with frame adverbials (see Sections 2.4 and 3.8.3).

However, the temporal units measured on continuous scales may be discretised (see Sections 2.6.2 and 3.6). We have seen that *po-*, *za-*, and *prze-* are particularly specialised as discretisers.

Durative temporalisation of perfectives is thus possible mainly with *po-*.DELIM meaning:

(64) a. *Poczytałam* *dwie* *godziny.*
DELIM.PFV.read.~~IPFV~~.PST.F.1SG two.ACC hour.PL.ACC
'I was reading a bit for two hours.'

b. * *Pozbierałam* *talerze* *dwie*
DISTR.PFV.pick.~~IPFV~~.PST.F.1SG plate.PL.ACC two.ACC
godziny.
hour.PL.ACC
('I was collecting the plates for two hours.')

c. * *Pojechałam* *do domu* *dwie* *godziny.*
along.PFV.drive.~~IPFV~~.PST.F.1SG to house two.ACC hour.PL.ACC
('I was driving home for two hours.')

d. ? *Pokochałam* *go* *dwie* *godziny*
DELIM.PFV.love.~~IPFV~~.PST.F.1SG he.ACC two.ACC hour.PL.ACC
'I loved him a bit for two hours. (inchoative interpretation impossible)'

e. * *Pokochałam* *go* *dwie* *godziny*
INCH.PFV.love.~~IPFV~~.PST.F.1SG he.ACC two.ACC hour.PL.ACC
('I fall in love with him for two hours' inchoative interpretation impossible)

Some prefixed verbs which encode the meaning ‘spend some time’ allow durative temporalisation; in these cases, the unprefixed stem must be imperfective. Verbs with the prefix *za-* such as *zabawić* ‘stay for a while’, *zaczekać* ‘wait for somebody or something’, verbs with prefixes *od-*, *prze-*, *wy-* (*od-siedzieć*, *wy-siedzieć*, *prze-siedzieć* (all meaning ‘sit through’) belong to this group.

(65) *Za-bawiłam tam dłużej, niż planowałam.*
behind.PFV-play.~~IPFV~~.PST.F.1SG there longer than plan.IPFV.PST.F.1SG
‘I stayed there longer than I had planned.’

The prefix *prze-* always requires an accusative durative adverbial, which is optional in the case of the prefix *wy-*:

(66) a. *Przesiedziałam na zebraniu cały dzień.*
PERDUR.PFV.sit.~~IPFV~~.PST.F.1SG on meeting.LOC all.ACC day.ACC
‘I spent the whole day on the meeting.’

b. *Przesiedziałam na zebraniu dwie godziny.*
PERDUR.PFV.sit.~~IPFV~~.PST.F.1SG on meeting.LOC two.ACC hours.PL.ACC
‘I spent two hours on the meeting.’

(67) a. *Ledwo wysiedziałam na zebraniu.*
barely out.PFV.sit.~~IPFV~~.PST.F.1SG on meeting.LOC
‘I could barely stand the meeting.’

b. *Wysiedziałam na zebraniu godzinę.*
out.PFV.sit.~~IPFV~~.PST.F.1SG on meeting.LOC hour.ACC
‘I was sitting at the meeting for an hour (and I could not anymore).’

Frame adverbials

The construction *w* + ACC belongs to the *frame adverbials* mentioned in Section 2.4. In the first place, it is possible in the case of perfectives whose meaning is compositional, in the sense that, apart from situation S_1 to which the verb refers,

the verbal meaning implies the existence of some preceding situation S_2 which could be expressed with an imperfective verb. For example the verb *przeczytać* 'to read through' implies a preceding situation referred to by the imperfective verb *czytać* 'to read', and the verb *znaleźć* 'to find' implies an existence of a preceding situation referred to by the imperfective verb *szukać* 'to look for'. The time referred to by the frame adverbial is the maximal interval from the beginning of S_2 to the end of S_1:

(68) *Przeczytałam tę książkę w miesiąc.*
through.PFV.read.~~IPFV~~.PST.F.1SG this.ACC book.ACC in month.ACC
'It took me one month to have read this book.'

Thus, one could not say:

(69) ? *W minutę znalazłam na ulicy złoty pierścionek.*
in minute.ACC find.PST.F.1SG on street.LOC golden.ACC ring.ACC
('I found a gold ring on the street in a minute')

if one meant that finding the ring was accidental, but it would be correct if finding was the result of searching.

Frame adverbials are thus incompatible with semelfactives and perfectives with delimitative or perdurative meaning:

(70) * *Poczytałam tę książkę w godzinę.*
DELIM.PFV.read.~~IPFV~~.PST.F.1SG this book in hour.ACC
('I read this book in an hour.')

(71) * *Kopnęłam piłkę w sekundę.*
kick.PFV.SEMEL.PST.F.1SG ball in second.ACC
('I kicked this ball in a second.')

Neither they can combine with the verbs of direct movement and prefix *po-*:

(72) * *Pojechałam do domu w dwie*
along.PFV.drive.~~IPFV~~.PST.F.1SG to house.GEN in two.ACC
godziny.
hour.PL.ACC
('I drove home in two hours.')

Frame adverbials may combine with imperfectives, when the meaning of the clause expresses the result rather than the duration:

(73) *W kilka minut była gotowa do drogi.*
in few minute.PL.ACC COP.IPFV.PST.F.3SG ready to way.GEN
'In a couple of minutes she was ready to go.' (Przepiórkowski et al. 2012)

Expression of degree in the form of adverbial phrase with temporal meaning

Some adverbial phrases have temporal meaning, but they function as expressions of degree (see Section 2.4). In Polish these take the form of the prepositional phrase *na* + ACC and are possible with both imperfectives (74) and perfectives (75).

(74) *Wyjeżdżam na dwa tygodnie.*
out.~~PFV~~.drive.IPFV.1SG on two.ACC week.PL.ACC
'I am leaving for two weeks.'

(75) *Wyjechałam na dwa tygodnie*
out.PFV.drive.~~IPFV~~.PST.F.1SG on two.ACC weeks.PL.ACC
'I (have) left for two weeks.'

3.8.4 Pluractionality

Temporal quantification

Both types of PVA appear in existentially quantified sentences. In a specifically quantified temporal context the usage of PVA can be directly translated on a scale between the discrete and the continuous.

(76) a. *Wczoraj na kolację zjadłam jabłko.*
yeasterday for dinner.ACC PFV.eat.~~IPFV~~.PST.F.1SG apple.ACC
'Yesterday, I ate an apple for dinner.'

b. *Wczoraj na kolację jadłam jabłko.*
yeasterday for dinner.ACC eat.IPFV.PST.F.1SG apple.ACC

‘Yesterday, I was eating apple for dinner.’

Existential quantification often applies to narrative, so the assignability of TR is restricted by taxis and durative temporalisation.

In terms of unspecific quantification, both aspects are possible too, but it is most likely that the IPFV is preferred, because the TSIT is assigned to a TR of unknown scalar properties.

(77) *Jadłam kiedyś homara.*
eat.IPFV.PST.F.1SG once lobster.ACC
‘I ate lobster once.’

(78) *Zjadłam kiedyś homara.*
froma.PFV.eat.~~IPFV~~.PST.F.1SG once lobster.ACC
‘I ate (a) lobster once.’

As Chapter 2 showed, sentences quantified universally are continuously valid in some part of timespace, but they do not represent occurrences of situations, usually because some element of this situation refers to a type. Assignment to the temporal units measurable on discrete scales is excluded, and therefore, they are formulated mostly with imperfectives, because perfectives are used for the TSIT assigned to a TR posterior to the TU:

(79) *Ziemia kręci się wokół Słońca.*
Earth turn.IPFV REFL around Sun.GEN
‘The Earth turns around the Sun.’

(80) *Ziemia zakręci się wokół Słońca.*
Earth behind.PFV.turn.~~IPFV~~ REFL around Sun.ACC
‘The Earth will turn a turn around the Sun’

(81) *Maria ładnie tańczy.*
Maria nicely dance.IPFV
‘Maria dances nicely.’ (in the sense: ‘If it happens that Maria is dancing, she is doing it nicely.’)

(82) *Maria ładnie zatańczy.*
Maria nicely behind.PFV.dance.~~IPFV~~
'Maria will dance nicely.' (e.g. tomorrow in the competition)

(83) *Diplodoki jadły liście.*
diplodocus.PL eat.IPFV.PST.NVIR.3.PL leaf.PL.ACC
'Diplodocuses ate leaves.' (in the sense: the species Diplodocus which once inhabited the Earth)

(84) *Diplodoki zjadły liście.*
diplodocus.PL from.PFV.eat.~~IPFV~~.PST.NVIR.3.PL leaf.PL.ACC
'Diplodocuses ate leaves.' (e.g. for lunch)

Statements similar to (85) pose one exception to this rule:

(85) *Lekarz dostanie pracę nawet w najdalszym zakątku świata*
doctor to.PFV.stand work.ACC even in furthest.LOC corner.LOC world.GEN
'A doctor would get a job even in the farthest corner of the world' (S1448)

Although the perfective verb appears in the Non-past tense, the sentence does not have future reference, but is rather an implicature: 'if a doctor searches for a job, he can get it anywhere', or 'the profession of medical doctor is needed so much that a doctor can get a job anywhere in the world'.

Quantification over referents

Polish does not overtly mark the type of referent quantification, for example, with a system of articles. Therefore, referents are quantified either within pragmatics or are separately marked in the nominal phrase by determiners, quantifiers, or numerals. The level of familiarity is purely pragmatic information, which does not relate directly to the choice of PVA in sentence.

However, the Accusative/Partitive alternation in Polish appears as the possibility of using a Genitive nominal object. It refers to an unspecified quantity. It is possible mainly with uncountable (mass) nouns and in the plural form in the

case of referents which could be considered uncountable. Partitive is usually associated with a part–whole relation, and, as the next chapter shows, in Finnish this case marking alternation is considered an aspectual correlate. That is, the partitive meaning is associated with IPFV and the holistic with PFV. In Polish, use of the Genitive-Partitive object undergoes some aspectual restrictions, but contrary to the intuitively assumed correspondence. Wierzbicka (1967: 2238) observes that a bare Genitive-Partitive object may only occur concurrently with PFV as in (86). However, it is a rarely noted fact that both aspectual values are possible when the object is additionally characterised as in (87).

(86) a. *On kupował *chleba /chleb.*
he buy.IPFV.PST bread.GEN bread.ACC
'He was buying bread.'

b. *On kupił chleba /chleb.*
he buy.PFV.PST bread.GEN bread.ACC
'He bought bread.'

(87) a. *On kupował chleba /chleb* ***na śniadanie***.
he buy.IPFV.PST bread.GEN bread.ACC for breakfast.ACC
'He was buying /bought bread for breakfast.'

b. *On kupił chleba /chleb* ***na śniadanie***.
he buy.PFV.PST bread.GEN bread.ACC for breakfast.ACC
'He bought bread for breakfast.'

Types of frequency

In Chapter 2, I distinguished three types of frequency: summaric, specific cycle and unspecific cycles.

The habitual imperfectives formed from simplex imperfectives are specialised in marking unspecific cycles:

(88) a. *Siad-yw-ał w kawiarni przy swoim stoliku*
sit-IPFV.HAB-PST in cafe.LOC at own.LOC table.LOC

'He used to sit at his (own) table in the cafe.'[12]

b. *Po wojnie pis-yw-ał do Szpilek*
after war write-IPFV.HAB-PST to Szpilki.GEN
'After the war he used to publish in Szpilki.'[13]

The summaric type of frequency and specific cycle must be introduced lexically, but it allows more variation in aspectual values, so both aspectual values may appear.

(89) *Spotykałam się z nim* ***trzy razy***.
meet.IPFV.PST.F.1SG REFL with him.INS three time.PL
'I met with him three times.'

(90) *Spotkałam się z nim* ***trzy razy***.
meet.PFV.PST.F.1SG REFL withhim.INS three time.PL
'I met with him three times.'

In specific cycles IPFV is often the only choice in the past temporal reference, but not in the future.

(91) a. *Brałem to lekarstwo co dwie*
take.IPFV.PST.M.1SG this.ACC medicine.ACC every two.ACC
godziny.
hour.PL.ACC
'I took this medicine every two hours.'

b. * *Wziąłem to lekarstwo co dwie*
take.PFV.PST.M.1SG this.ACC medicine.ACC every two.ACC
godziny.
hour.PL.ACC
('I took this medicine every two hours.')

[12] http://kobieta.gazeta.pl/kobieta/56,125640,11660046,kawiarnie-kiedys-i-dzisiaj-czy-bardzo-sie-zmienily.html
[13] https://www.tygodnikprzeglad.pl/kern-zegna-sie-przekrojem/

(92) *Wezmę to lekarstwo co dwie godziny.*
take.PFV.1SG this.ACC medicine.ACC every two.ACC hour.PL.ACC
'I will take this medicine every two hours.'

(93) *Będę brać to lekarstwo co dwie*
AUX.1SG take.IPFV.INF this.ACC medicine.ACC every two.ACC
godziny.
hour.PL.ACC
'I will take this medicine every two hours.'

Such theoretical observations are supported by empirical studies. Stawnicka (2007) examines the relationship between the type of frequency marker and PVA based on the corpus material. In all three types IPFV dominates. The only exception is the summaric, specific frequency, as in the examples above. In that case, the distribution between PFV and IPFV is balanced. This is in line with the discrete assignability of the TSIT of perfectives. The specific summaric frequency is discrete, thus compatible with PFV.

3.9 Conclusions

This chapter has shown that the semantics of PVA can be described based on the distinction between temporal units measured on discrete and continuous scales, and it refers to the scalar compatibility between the TSIT and TR to which TSIT can be assigned expressed in through PVA. In other words, PVA specifies what kind of TRs TSIT can be assigned to.

PVA interacts with other temporal elements of Polish. It supports the temporal-deictic function in interaction with tense. With tense and lexical markers, it also has an ordering function, indicating simultaneity (mainly IPFV) and non-simultaneity (PFV).

PVA is part of the temporal system of Polish, but PVA markers often contribute to the lexical meaning of the verb. Prefixes, which are markers of PFV in the absence of a secondary imperfective thematic suffix, developed mostly from prepositions, and they enrich the meaning of the base with the spatial parameter.

Additionally, some prefixes developed narrow, mostly temporal meanings called Aktionsart. Durative temporalisation is an operation applicable to the TSIT of situations denoted in imperfectives. Durative temporalisation of the TSIT in

perfectives is restricted to verbs which primarily encode the meaning ‘spend some time’. The prefixes *po-*, *za-*, *prze-* (and, to a lesser extent, some other prefixes) may act as discretisers. In particular with DELIM.*po-*, TR is a discretised temporal interval, while *za-* discretises mostly thanks to the semantically implicit existence of inherent-scale argument.

In the context of past temporal reference, the distribution of frequency type is unrestricted in IPFV, whereas PFV can appear in the context of a discrete, limited type of frequency.

On the level of discourse, PVA interacts with the temporal quantification of the sentence. While both aspects are used in existential quantification, IPFV applies typically to universally quantified sentences. IPFV is also preferred when the TSIT has a discontinuous temporal structure. PFV is used mainly when the number of TRs is finite and known.

Chapter 4

Temporality in Finnish

4.1 Introduction

Temporality in Finnish indicative clauses is expressed through various tools of both temporal localising (Section 4.4) and durative temporalisation (Section 4.5). However, the grammatical category of aspect, in the sense of the Polish opposition between PFV and IPFV, is not a grammatical category of Finnish verbs. Neither has Finnish developed the system of articles or quantifiers which would enable any particular type of quantification over referents.

However, semantic-syntactic relations can be expressed in Finnish by means of fifteen case affixes, a relatively big number in comparison to Indo-European languages. Bearing in mind that languages can map parameters of TSIT to arguments (Section 2.6.4), I take a closer look at whether and how any properties of TSIT can be expressed in the case marking of certain arguments (Section 4.3). This would compensate for the lack of spatial prefixes applied directly to the verbal lexeme, as in Slavic, German or even Hungarian, which belongs to the same language group as Finnish.

This chapter is constructed analogically to Chapters 2 and 3, but since I discuss the contribution of the nominal phrase to the Finnish temporal system, I also sketch the case marking system, and describe it in the context of argument marking.

4.2 Basic remarks on the Finnish grammatical system

4.2.1 Finnish verbs

Morphological structure

Finnish verbal morphology can be described according to the following template:

[(compound element) + root + (max. 3 suffixes) VERBAL STEM] + agglutinative endings

Most roots end, as in Polish, with a vowel, but it is not a strict rule. Some roots are free morphs as in the verb *nolo-tta-a* 'to make somebody embarassed', where *nolo* means 'embarassed', but but in some verbs as for example in *odotta-a* 'to wait' no free morph root can be identified.

The root may be followed by a maximum of three suffixes belonging to different semantic-syntactic types (Section 4.2.1. The most common suffixes in Finnish are: *-U-*,[1] *-O-*, *-i-*, *-AA-*, and *-itse-* (Hakulinen et al. 2004: §149).[2]

The core of the verbal stem is therefore the root and suffixes. Most verbs have a vowel-ended, strong (94a) and weak stems (94b). This distinction refers to the qualitative and quantitative consonant gradation in the verbal stem. Some verbs have an additional consonant-ended stem, used for instance to form the First Infinitive (henceforth INF1) (94c), the third person singular, and the second person plural of the imperative:

(94) *tehdä* 'to do'

a. *teke-e*
do-3SG
'he/she/it does'

b. *tee-n*
do-1SG
'I do'

[1]The uppercase vowel in Finnish morphemes stands for variants of vowel harmony: U=u/y, O=o/ö, A=a/ä

[2]See Hakulinen et al. (2004: §304,§306) for an exhaustive list and some rules for combining suffixes.

c. *teh-dä*
do-INF1
'to do'

Henceforth, I refer mainly to Finnish verbs in their INF1 form. However, when discussing verbal suffixes, the use of strong stems ending in a hyphen is the better choice, since in INF1 the derivational suffix is sometimes reduced.

Verbal suffixes

Finnish verbs are formed mainly by suffixation. Additionally to plain *verbalisers* (Laakso 1997) – suffixes whose primary function is marking belonging to a verbal category – three types of suffix are listed, as they carry additional information as to syntactic constraints or the inner temporal structure of the situation referred to by a verb:

1. transitivisers (fin. *muuttamisjohdokset*)[3]
2. detransitivisers (fin *muuttumisjohdokset*)
3. modifiers (fin. *muuntelujohdos*)

I now discuss these suffixes briefly, whereas Table A.1 in Appendix A presents the main derivational verbal classes in detail.

Detransitivisers Detransitivisers relate to valency reduction and reflexivity,[4] which as described by Hakulinen et al. (2004: §303, §333), often belongs to the verbal stem in Finnish. The most important morphemes in this group are: *-UtU- pese-yty-* 'to wash', *-VntU- kerä-änty-* 'to gather, to pile up, to accumulate', *-U- siirt-y-* 'to move, to be transferred', *-tU- nuor-tu-* 'to become younger', *-ne- suure-ne-* 'to widen'.

[3]The Finnish terms in parenthesis refer to terminology used in Hakulinen et al. (2004).

[4]Nevertheless, reflexivity can be also expressed in the direct object:

(1) *...käärme tosiaan syö itse-ä-än*
snake really eat [self-PAR-3 OBJECT]
'... the snake is really eating itself.'[5]

Functions of detransitivisers can be described within the same frame, as I used for the Polish reflexive marker *się* (see Section 3.3). However, lexical reflexives are not possible in Finnish. Besides genuine reflexives (e.g. *peseyt-y-* 'to wash yourself') and reciprocals, detransitivisers fulfil a decausative function as in (2) and serve as passive reflexives (3):

(2) *Köysi kirist-y-y.*
rope tighten-REFL-3SG
'The rope tightens itself.'

(3) *Automatisoitu ovi ava-utu-u helposti nappia painamalla.*
automatic door open-REFL-3SG easily button.PAR press.INF3.ADE
'Automatic doors open easily by pressing the button.'[6]

The morpheme *-ne-* attached to the adjectival root forms a change in quality described by the root: *pime-ne-* 'to darken', *kove-ne-* 'to harden'. I call such verbs translatives, similarly to the Polish group of verbs.[7]

Transitivisers Valency extension refers mainly to causativity and it is often marked explicitly for example with morphemes: *-ttA- syö-ttä-* 'to feed' from *syö-* 'to eat', *-tA- havainnollis-ta-* 'to illustrate', morpheme *-Oi- satul-oi-* 'to saddle' and morpheme *-itse- kalk-itse-* 'to whitewash'. It is possible to distinguish semantically uniform, predominantly denominal derivative classes of verbs containing a causative morpheme, but these do not influence the syntactic properties of verbs.

Temporal structure modifiers Temporal structure modifiers are suffixes which introduce a change in the inner temporal structure of the situation which is not possible to conclude from a corresponding simpler verbal stem.

FREQUENTATIVE morphemes: *-le- odotte-le-* ('to wait'), *-hti- hyppele-hti-* ('to bob, to jog, to jump'), *-i- ysk-i-* ('to cough'), *-o- huit-o-* ('to lay about'), *-ise- vär-ise-* ('to shake, to shiver') mark prototypically situations of which perception is longer than the situation described by the correlate verb (4) or as a repeating notion (5):

[6] `http://www.jkovet.fi/talvikamppanja-tasmalaake-pakkaseen/`

[7] Although *essive* and *translative* are broadly used as names for cases, in the Uralic literature they also designate semantic verb groups, for example, in Kiefer & Laakso (2014). In the typological literature the term *inchoative* is sometimes used, for verbs morphologically marking the meaning 'to become like Adj' (cf. Haspelmath & Sims 2010).

(4) a. *Grillaa-n makkaraa vain kerran viikossa.*
grill-1SG sausage.PAR only once week.INE
'I grill sausage only once a week.'[8]

b. *Jos grillailemme pimeinä syysiltoina*, on myös ampiaisia paikalla.

jos grilla-ile-mme pimeinä syysiltoina
if grill-FREQ-1PL dark.PL.ESS autumn.night.PL.ESS
'*If we are grilling on dark autumn nights*, wasps appear.'[9]

(5) a. *Laiva heil-aht-i vaarallisesti.*
boat swing-MOM-SPST dangerously
'The boat rocked dangerously.'

b. *Faktantarkastajien valhemittari heil-aht-el-i*
fact.checker.PL.GEN lie.detector swing-MOM-FREQ-SPST
Trumpin ja Clintonin vaaliväittelyn aikana.
Trump.GEN and Clinton.GEN election.debate.GEN time.ESS
'The lie detector of fact checkers was swinging to and fro during the presidential debate of Trump and Clinton.'[10]

Frequentatives may also refer to irregular, non-serious or accidental situation:

(6) a. *Hän luk-i lehteä.*
(s)he read-SPST newspaper.PAR.
'(S)he was reading a newspaper.'

b. *Hän lue-skel-i lehteä.*
(s)he read-FREQ-SPST newspaper.PAR.
'(S)he was browsing a newspaper.'

[8] http://syohyvaa.fi/stressivapaa-vinkkilista-lomakilojen-kontrollointiin/
[9] http://yle.fi/uutiset/3-6833176
[10] http://yle.fi/uutiset/3-9219619

The MOMENTANEOUS morphemes: *-AhtA-* (mostly intransitive) *ist-ahta-* 'to sit down', *-Aise-* (mostly transitive) *kys-äise-* 'to ask in passing' are used in the first place for denoting situations perceived as lasting for a very short period of time, or similarly to Polish semelfactives, to refer to the smallest quantum of action as in (5a) above.

The suffixes *-i-*, *-ksi-* may mark a continuous situation in comparison to some verbal correlate, for example, *hypp-i-* 'to jump for a longer time' – *hypp-ää-* 'to jump once', *lykk-i-* 'to push for a longer time' – *lykk-ää-* 'to push once' (Laakso 1997: 294). The semantic difference between such *continuatives* and frequentatives is hard to grasp. Therefore, in the current study I do not distinguish between these two types.

Expressing attitude The morpheme *-ksU-* appears in mostly deadjectival, transitive verbs and it reflects the meaning of approach or attitude of the subject towards the object: *hyvä-ksy-* 'to approve', *pahe-ksu-* 'to disprove, deprecate' (Hakulinen et al. 2004: 350).

Essives Denominal, mostly intransitive verbs are often derived with suffixes homographic to transitivisers as in: *kumi-tta-* 'to erase with an eraser', *skype-ttä-* 'to chat via Skype' *käännyk-öi-* 'to call with a mobile phone', *tango-a-* ('to dance tango'), *sairas-ta-* 'to be ill'. The morpheme *-ile-* appears in that function as in *pyörä-ile-* 'to bike'.

Loan verbs As noted in Kiefer & Laakso (2014), *-AA-* is nowadays the most frequent way of forming new verbs from monosyllabic loanwords, as in *buukk-aa-* 'to book'. Multisyllabic loanwords are adapted with the morpheme *-Oi-* as in *absorb-oi-* 'to absorb'. The morpheme *-itse-* had been used in the older germaninc loans *val-itse-* 'to choose', *tuom-itse-* 'to judge' (Laakso 1997: 275).

Free ad-verbals

Genuine verbal prefixation is atypical in Uralic languages. Nevertheless, morphologically or syntactically complex verbal lexical units are formed in Finnish with adpositions, adverbs and particles, which I call FREE AD-VERBALS.[11]

[11] I underline their syntactic freeness in contrast to compound lexemes which originate from non-finite forms and usually do not represent the full active, finite paradigm. This class of compounds has been thoroughly described by Kolehmainen (2005).

Free ad-verbals are usually adverbs, particles or adpositions, often in the concrete case forms (see Section 4.2.2). Kolehmainen (2005: 136) gives the following list of free ad-verbals:

1. Unsegmentable: *ensi* 'first', *vasta* 'recently, fresh', *viime* 'last'
2. Adverbial suffix: *uudesti* 'again'
3. Prolative: *ohitse* 'past', *ylitse* 'above'
4. Translative: *luokse* 'towards', *taakse* 'behind', *viimeksi* 'last'
5. Instructive: *ennen* 'before', *ensin* 'first', *kesken* 'middle', *takaisin* 'back', *väärin* 'wrong'
6. Illative: *esiin* 'out', *eteen* 'in the front', *jälkeen* 'after', *kokoon* 'together', *mukaan* 'with', *oheen* 'next to', *perään* 'after', *sisään* 'in/inside', *talteen* 'into safekeeping', *umpeen* 'closed, blocked', *väliin* 'between', *varteen* 'upon', *vastaan* 'against', *yhteen* 'together'
7. Allative: *alle* 'under', *edelleen* 'further', *esille* 'in front of', *jälleen* 'again, new', *koolle* 'together', *päälle* 'on top of'
8. Old Lative forms: *alas* 'under', *ali* 'under', *halki* 'atwain', *ilmi* 'revealed', *irti* 'away', *julki* 'public', *juuri* 'just, current', *kiinni* 'closed', *läpi* 'through', *liki* 'nearly', *ohi* 'beside, in passing, over', *poikki* 'atwain', *pois* 'away', *rikki* 'broken', *ulos* 'out, outwards', *ylen* 'over', *yli* 'over', *ylös* 'high, upwards', *ympäri* 'around'
9. Adessive: *edellä* 'in front', *sisällä* 'in', *välillä* 'between'
10. Inessive: *edessä* 'in front', *koossa* 'together', *perässä* 'after, behind', *poissa* 'away'
11. Old Locative: *alla* 'under', *läsnä* 'present', *mukana* 'with', *takana* 'behind', *yllä* 'above'
12. Ablative: *päältä* 'from, from the top'
13. Elative: *edestä* 'from in front of'
14. Old Separative forms: *luotaan* 'away from', *myötä* 'with, along', *takaa* 'from behind'

A free ad-verbal may be attached to the beginning of the verb, or it may stand alone in the sentence, as below:

(7) *irti+sanoa*

a. Koeajalla Kärpäsen koulussa ollut kieltenopettaja kertoo lehdelle, *että hänet on irtisanottu*.

että hänet on ***irtisanottu***
that (s)he.ACC AUX off.tell.PSTPP
'The language teacher who has been on the trial period in the school in Kärpänen tells the newspaper *that (s)he has been given her notice*.'[12]

b. Työtodistukseen on pantava, *että hänet on sanottu irti.*

että hänet on ***sano-ttu irti***
that (s)he.ACC AUX tell.PSTPP off
'It must be written in the reference, that (s)he has been given her notice.'[13]

(8) *vastaan+ottaa*

a. *SAS* ***vastaanotti*** *ensimmäisen A320neo-lentokoneensa.*
SAS towards.take.SPST first.GEN A320neo.aeroplane.POSS3
'SAS received its first A320neo aeroplane.'[14]

b. *Norwegian* ***otti vastaan*** *ensimmäisen Airbus 320neo-lentokoneen*
Norwegian take.SPST towards first.GEN Airbus

A320neo.aeroplane.GEN
'Norwegian received its first Airbus 320neo aeroplane.'[15]

[12] http://yle.fi/uutiset/3-9271936
[13] http://www.helsinki.fi/kansalaismuisti/savonlinna/rouva.htm
[14] http://www.is.fi/matkat/art-2000001939345.html
[15] http://www.tekniikkatalous.fi/talous_uutiset/liikenne/norwegian-otti-vastaan-ensimmaisen-airbus-320neo-lentokoneen-tilannut-yhteensa-70-6606071'

Ad-verbals should not be confused with fixed or idiomatic expressions, mostly formed with nominal parts of speech:

(9) a. *saa-da* *aika-an*
get-INF1 time-ILL
'to organise'

b. *käy-dä* *käs-i-ksi*
walk-INF1 hand-PL-TRANS
'to attack'

The relation between ad-verbals and temporality is discussed further in Section 4.3.4 in the context of resultative constructions.

4.2.2 The case inventory

The central nominal category is case. Following the terminological recommendations of Haspelmath (2009), cases can be divided into abstract cases associated with core syntactic relations and concrete cases related mostly to semantic ones. The latter group can be additionally divided into spatial and nonspatial cases. This classification fits roughly to the Finnish case system (see Table 4.1 below), although a clear-cut distinction is not always possible.

Case	SG	PL	Typical function
Abstract cases			
Nominative	kissa	kissa-t	subject, object
Genitive	kissa-n	kisso-je-n	possessor, object
Accusative	minu-t	meidät	object, possessed
Partitive	kissa-a	kisso-j-a	object, the partial nature of the referent of the noun
State cases			
Essive	kissa-na	kisso-i-na	temporary, changeable state
Translative	kissa-ksi	kisso-i-ksi	change of state
Spatial cases			
Allative	kissa-lle	kisso-i-lle	movement 'on, onto'
Adessive	kissa-lla	kisso-i-lla	location 'on', possessor
Ablative	kissa-lta	kisso-i-lta	movement 'from, away'
Illative	kissa-an	kisso-i-hin	movement 'in, into'
Inessive	kissa-ssa	kisso-i-ssa	location 'in, inside'
Elative	kissa-sta	kisso-i-sta	movement 'out', source
Concrete non-spatial cases			
Abessive	kissa-tta	kisso-i-tta	'without'
Comitative	-	kisso-ine-en	companion
Instructive	jalan	jaloin	instrument
Prolative	-	puhelim-i-tse	chanel, instrument

Table 4.1: Finnish cases.

Note: For all cases but Accusative, Instructive and Prolative the paradigm of the lexeme *kissa* 'cat' is given. The Accusative applies only to personal pronouns, here: *minä* 'me' and *me* 'we'. The example for the Instructive means 'on foot', for example, *to go on foot*, and 'with feet', for example, *paljain jaloin* 'barefoot', while for the Prolative 'on the phone', 'by phone'.

Abstract cases

Abstract cases are typically used for subject, object and predicative. They often appear in the constructions characterising the relation of possession.

The subject appears mainly in the Nominative as in (10a). The subject of indefinite quantity, in particular in existential sentences, appears in the Partitive as in (10b):

(10) a. *Lapse-t juokse-vat leikkikentä-llä.*
child-PLNOM run-3PL playground-ADE
'(The) children are running on the playground.' (Huumo et al. 2017)

b. *Leikkikentä-llä juokse-e laps-i-a.*
playground-ADE run-3SG children-PL-PAR
'There are children running on the playground.' (Huumo et al. 2017)

The nominal object can be marked with all four abstract case suffixes. This property of Finnish belongs to the phenomenon called DIFFERENTIAL OBJECT MARKING (DOM) which was introduced to linguistics by Bossong (1985). While the term PARTITIVE OBJECT refers to the direct object in the Partitive, Nominative, Genitive, and Accusative are usually treated as serving the same semantic function and they are labelled under a common term, TOTAL OBJECT. Nonetheless, the Total object cases are not interchangeable: the Accusative is used only with personal pronouns, while the Nominative applies to plural forms and singular forms in impersonal sentences. The Total-Partitive opposition and DOM is discussed in more detail in Section 4.3 as it is important in specifying the properties of TSIT in Finnish.

All abstract cases but the Genitive are used to mark the possessed argument in the possessive sentence, with the 3SG form of the verb *olla* 'to be' and the first position possessor argument in the form of the Adessive:

(11) a. *Minu-lla on lapsi.*
me-ADE be.3SG child
'I have a child.'

b. *Minu-lla on aika-a.*
me-ADE be.3SG time-PAR

‘I have time.’

c. *Toiseksi minu-lla on epäilyks-i-ä komission roolista*
secondly me-ADE be.3SG doubt-PL-PAR commision.GEN role.ELA
tässä asiassa-
this.INE think.INE
‘Secondly, I have some doubts about the commision’s role in this matter.’ (Borin et al. 2012)

d. *Sen lisäksi minu-lla on valtuude-t*
this.GEN extra.TRANS me-ADE be.3SG authority-PL.NOM
neuvotella maatalousasioista esimerkiksi WTO:ssa.
negotiate.INF1 agrar.matter.PL.ELA example.TRANS WTO.INE
‘Additionally, I have the powers to negotiate on agriculture matters, for example in the WTO.’ (Borin et al. 2012)

e. *Minu-lla on sinu-t.*
me-ADE be.3SG you-ACC
‘I have you.’

The Genitive is used to mark the possessor in the noun phrase:

(12) *minu-n kynä-ni*
me-GEN pen-POSS.1SG
‘my pen’

In neccessive constructions the Genitive is used to mark the argument to which the necessary action applies:

(13) a. *Minu-n täytyy mennä.*
I-GEN must.3SG go.INF1
‘I must go.’

b. *Minu-n ei tarvitse mennä.*
I-GEN NEG need go.INF1
‘I do not need to go.’

State cases

The Translative and the Essive are usually classified by scholars either to spatial or to abstract cases. Tommola (1986), following Siro (1964), classifies them to spatial cases. Hakulinen et al. (2004) calls them *abstract spatial* cases, Hynönen (2017), following Haspelmath (2009), classifies them as *abstract* cases, although at least the Essive has a clear old Uralic locative origin. The Translative and the Essive are used mainly to express non-permanent state relations, but they also have temporal (see Sections 4.4.1 and 4.5) and locational functions.

The Essive is a marker of non-permanent states or circumstances, for example, as a predicative in (14b) or oblique argument in (15):

(14) a. *Hän on sairas /opettaja.*
(s)he COP ill /teacher
'(S)he is ill /(professionally) a teacher' (Hynönen 2017: 30)

b. *Hän on sairaa-na /opettaja-na.*
(s)he COP ill-ESS /teacher-ESS
'(S)he is (temporarily) ill /(working as) a teacher.'(Hynönen 2017: 30)

(15) *Pidän Hannaa hyvä-nä ystävä-nä.*
regard.1SG Hanna.PAR good-ESS friend-ESS
'I regard Hanna as a good friend.'

The Translative,in contrast, marks the result of change from one state to another, thus end state:

(16) *Hän jää leske-ksi.*
(s)he remain widow-TRANS
'(S)he will become a widow(er).'

(17) *Kesä muuttui kylmä-ksi.*
summer change.REFL.SPST cold-TRANS
'The summer became cold.'

Spatial cases

The basic function of the six spatial cases is marking spatial relations. The Inessive expresses the static position *in*, the Adessive the static position *on*:

(18) a. *Kirje on laatiko-ssa.*
letter be.3SG box-INE
'The letter is in the box.'

b. *Kirje on pöydä-llä.*
letter be.3SG table-ADE
'The letter is on the table.'

The Illative and the Allative mark movement towards, while the Ellative and the Ablative mark movement away:

(19) a. *Hanna meni kirjasto-sta koti-in.*
Hanna go.SPST library-ELA home-ILL
'Hanna walked home from the library.'

b. *Hanna meni tori-lta /isä-ltä juna-asema-lle /mummo-lle.*
Hanna go.SPST market-ABL /father-ABL train.station-ALL /granny-ALL
'Hanna walked from the market /the father to the train station /the granny.'

Thus, two types of opposition can be distinguished in the spatial system. The first one, which concerns marking the direction of movement against lack of movement (or undirected movement). This opposition is called LATIVE – ESSIVE opposition. According to Kangasmaa-Minn (1984: 77) it defines the character of the situation; lative cases characterise dynamic situations, while essive cases define static ones.

The second opposition is formed by two series of local cases: the external formed by the Allative, the Adessive and the Ablative, and the internal formed from the Illative, the Inessive and the Elative (Huumo & Ojutkangas 2006: 12). The external cases are used to mark spatial relations involving animate landmarks, as shown in (19b).

Consequently, the owner is marked with the same series of cases. The static relation was shown above in (11), the cases marking movement are used to mark a change of owner:

(20) a. *Sain kirjan Hanna-lta.*
receive.SPST.1SG book.GEN Hanna-ABL
'I received the book from Hanna.'

b. *Annoin kirjan Hanna-lle.*
give.SPST.1SG book.GEN Hanna-ALL
'I gave the book to Hanna.'

The six cases have various non-spatial usages related to stimulus-experiencer meaning, instrumental, and whole-part usages. Due to lack of space they will not be discussed now, but the temporal usages are described in Sections 4.4.1 and 4.5, while their contribution to theme-to-event homomorphism is addressed in Section 4.3.

Concrete nonspatial cases

The concrete nonspatial cases are the least frequently used cases in modern Finnish. For most nouns they do not belong to the paradigm, and a competitive construction exists:

(21) a. *ystäv-i-ne-en /ystäv-i-e-nsä kanssa*
friend-PL-COM-POSS.3 /friend-PL-GEN-POSS.3 with
'with his friends'

b. *mu-i-n keino-i-n / mu-i-lla*
other-PL-INSTR method-PL-INSTR /other-PL-ADE method-PL-ADE
keino-i-lla

'with other methods'

Their contribution to temporality in Finnish is mainly in the form of temporal adverbial phrases.

4.3 Theme-to-event homomorphism

4.3.1 Transitivity

Having introduced the Finnish case system, I now discuss how it relates to TSIT in Finnish. The central notion here is the direct object and transitivity, with which I begin.

I define TRANSITIVE CONSTRUCTION in Finnish as a syntactic structure with an overt direct object.[16].

The object position can be filled with a nominal phrase, non-finite verbal clause, subordinate clause or direct speech. I now concentrate on the nominal object, of which case-marking has been seen as a central category correlating with Slavic verbal aspect. Such results were presented in the questionnaire-based cross-linguistic study by Dahl (1985), partly corpus-based, Finnish-Russian contrastive study of Tommola (1986), and Finnish-Czech contrastive study of Zmrzlíková (2009), as well as the monolingual corpus-based study of Askonen (2001). Theoretical descriptions by Hakulinen et al. (2004) include the case-marking opposition in the field of aspectual notions.

Finnish verbs rarely appear exclusively in transitive or intransitive constructions. The verb *kantaa* 'to carry' may appear without an object, although the most frequent meaning is modified here:

(22) a. *Lehmä kantaa vasikka-a mahassaan yhdeksän kuukautta.*
cow carry.3.SG [calf-PAR$_{OBJ}$] stomach.INE.POSS.3 nine month.PAR
'The cow carries the calf in her stomach for nine months.'[17]

b. *Lehmä kantaa noin yhdeksän kuukautta.*
cow carry.3.SG about nine month.PAR
'Cow's pregnancy lasts for about nine months.'[18]

[16]I prefer to discuss transitive constructions than transitive verbs, because nearly any Finnish verb can appear in a sentence with direct object.

[17]http://www.virtuaali.info/opetusmaatilat/index.php?tila_id=26I\&ohjemappiI\&kategoria_id=279I\&kortti=1348

[18]https://oikeuttaelaimille.fi/tuotantoelaimet/nauta-tuotantoelaimena

Leino (1991: 35) suggests that transitivity in Finnish should be treated as a polar opposition.[19] Basing on the distributional properties of verbs, the following syntactic types could be distinguished here:

- verbs which appear extremely rarely in transitive constructions, such as *tykätä* 'to like', mainly in resultative constructions (see Section 4.3.4):

(23) a. *Tykkään sinu-sta.*
like.1.SG you-ELA
'I like you.'

b. *Hän on tykännyt minu-t tervee-ksi.*
(s)he AUX like.PSTAP [me-ACC$_{OBJ}$] healthy-TRANS
'He liked me so I became healthy.'

- verbs which appear mostly in intransitive constructions, in transitive constructions only with an object of content (see Section 4.3.2), such as *elää* 'to live':

(24) a. *Anna eli kauan.*
Anna live.SPST long
'Anna lived long.'

b. *Anna elää yksinkertais-ta elämä-ä.*
Anna live.3SG [simple-PAR life-PAR$_{OBJ}$]
'Anna leads a simple life.'

- verbs which appear in both types of constructions, such as *kutoa* 'to knit':

(25) a. *Mummo kutoo sohvalla.*
Granny knit.3SG couch.ADE
'Granny is knitting on the couch.'

[19] This approach is close to, but not identical with, Hopper & Thompson (1980), but differs from Haspelmath (2015).

b. *Mummo kutoo sukk-i-a sohvalla.*
Granny knit.3SG [socks-PL-PAR$_{OBJ}$] couch.ADE
'Granny is knitting socks on the couch.'

- verbs which appear either in transitive constructions (26a) or require some other argument which changes the verb's lexical meaning (26b), such as *muuttaa* 'to move; to change; to amend':

(26) a. *Olen muuttanut mielipide-ttä-ni.*
AUX.1SG change.PSTAP [opinion-PAR-POSS.1SG$_{OBJ}$]
'I have changed my opinion.'

b. *Ensi vuonna muutan maa-lle.*
next year.ESS move.1SG country-ALL
'I will move to the countryside next year.'

- verbs which extremely rarely appear in the intransitive constructions, such as *ottaa* 'to take':

(27) *Otan ruoka-a jääkaapista.*
take.1SG [food-PAR$_{OBJ}$] fridge.ELA
'I will take some food from the fridge.'

(28) *Jos otan, en aja.*
if take.1SG NEG.1SG drive
'If I drink, I do not drive.'

4.3.2 Semantic types of the nominal object

The nominal object cannot be analysed as an aspectual marker without examining its semantic types, because the semantic type of an object determines whether and how the object can be quantified. There is no agreement among scholars as to how many semantic types can be assigned to the constructions fulfilling the structural criteria of the nominal object in Finnish. Leino (1991: 37) lists four main semantic

types, (Askonen 2001: 485) six. The tentative list below provides a typology of objects in Finnish based, on the one hand, on the semantic types applicable to the Finnish nominal object, and on the other hand, on such properties of an object as affectedness and providing some scale for measurement.[20] This list is revised in Chapter 7, based on the analysis of the empirical material.

1. argument-scale object: *patient* – undergoes a qualitative (29), quantitative (30) change:

   (29) *Kuorin* *omena-n.*
   peel.SPST.1SG apple-GEN
   'I peeled the apple.'

   (30) *Söin* *omena-n.*
   give.SPST.1SG apple-GEN
   'I ate the apple.'

2. path-scale object: *mover* – changes the location and/or the owner, but does not undergo a physical qualitative or quantitative change:

   (31) *Annan* *kirja-n* *Maria-lle.*
   give.1SG book-GEN Maria-ALL
   'I will give the book to Maria.'

3. inherent-scale object – this remains unaffected by the situation, but provides a property that could serve for measuring the situation:

   (32) *Luin* *lehde-n.*
   read.SPST.1SG newspaper-GEN
   'I read the newspaper.'

4. unaffected object:

   i) *positioned* – is located somewhere, or being someone's property, does not undergo a qualitative nor quantitative change:

[20] This description includes only the direct object, not the measure adverbials in the object cases which are morphologically identical with the nominal object. See Section 4.3.5.

(33) *Omistan talo-n Kotka-ssa.*
own.1SG house-GEN Kotka-INE
'I own a house in Kotka.'

ii) *stimulus* – is a source of some emotion, perception, or cognitive process:

(34) *Rakastan Hanna-a.*
love.1SG Hanna-PAR
'I love Hanna.'

iii) *experiencer* – experiences some emotion, feeling, or cognitive process:

(35) *Minu-a janottaa.*
me-PAR be.thirsty.3SG
'I am thirsty.'

iv) *reference* – this exists and remains unaffected by the situation, quantitatively, qualitatively, or spatially :

(36) *Hanna leikkii piilos-ta.*
Hanna play.3SG hide.and.seek-PAR
'Hanna plays/is playing hide-and-seek.'

Additionally to the above-mentioned roles, Finnish scholars (Askonen 2001; Hakulinen & Karlsson 1979; Hakulinen et al. 2004; Leino 1991) distinguish the OBJECT OF CONTENT which is redundant to the lexical meaning of a verb that normally does not appear in a transitive construction:

(37) *Anna elää yksinkertais-ta elämä-ä.*
Anna live.3SG simple-PAR life-PAR
'Anna leads a simple life.'

4.3.3 Total-partitive object opposition

As mentioned in Section 4.2.2, the Finnish object is marked with four abstract cases which form a binary opposition between Total and Partitive.

According to some authors (Groundstroem 1988; Hakulinen et al. 2004; Vilkuna 2000) the usage of the Total object boils down to three jointly applied, sufficient rules:

1. the sentence must be affirmative
2. the interpretation of the sentence must be bounded (the situation must be described as one unit which has a beginning, an end, or is punctual Hakulinen et al. 2004: 1498),
3. the object is quantitatively definite and the reference exhaustive.

These rules summarise many cases of DOM, but it is important to underline that the choice of object case depends on a wide range of factors. Here is a more detailed list of oppositions involved in DOM:

(38) Indefinite – definite quantity:

a. *Tämä laatikko sisältää vanha-t kirjee-ni.*
this box contain.3SG old-PL.NOM letters-POSS.1SG
'This box contains all of my old letters.'

b. *Tämä laatikko sisältää vanho-j-a kirje-i-tä-ni.*
this box contain.3SG old-PL-PAR letter-PL-PAR-POSS.1SG
'This box contains some of my old letters'

(39) Whole – part of the object:

a. *Näin kirkon kato-n puiden takaa.*
see.SPST.1SG church.GEN roof-GEN tree.PL.GEN behind
'I saw/was seeing the church roof behind the trees.'

b. *Näin kirkon katto-a puiden takaa.*
see.SPST.1SG church.GEN roof-PAR tree.PL.GEN behind
'I saw/was seeing a part of the church roof behind the trees.' (Tommola 1986: 140)

(40) Total – partial affectedness of the object:

a. *Söin omena-n.*
eat-SPST-1SG apple-GEN
'I ate the apple.'

b. *Söin omena-a.*
eat.SPST.1SG apple-PAR
'I ate/was eating an apple/some apple.'

(41) Gradability of the situation:

a. *Nostan sinu-t.*
raise.1SG you-ACC
'I (will) raise you (somewhere).'

b. *Nostan sinu-a.*
raise.1SG you-PAR
'I (will) raise you sa bit.'

(42) Lexical resultativity (the situation leads to some result):

a. *Metsästäjä ampui vahingossa lehmä-n.*
hunter shoot.SPST accident.INE cow-GEN
'The hunter shot a cow by accident.'

b. *Metsästäjä ampui vahingossa lehmä-ä.*
hunter shoot.SPST accident.INE cow-PAR
'The hunter shot (at) a cow by accident.' (Heinämäki 1984: 153)

(43) Affirmation versus negation:

a. *Rakensin talo-n.*
build.SPST.1SG house-GEN
'I built a house.'

b. *En rakentanut talo-a.*
NEG.1SG build.PSTAP house-PAR

‘I did not build a house.’

(44) Type versus token, which Tommola (1990: 360) calls non-specific versus specific end state:

a. *Mies vaihtoi paita-a.*
man change.SPST shirt-PAR
‘The man changed a shirt.’

b. *Mies vaihtoi puhtaa-n paida-n.*
man change.SPST clean-GEN shirt-GEN
‘The man put on the clean shirt.’ Tommola (1990: 360)

(45) Constant quantity over a continuous unit of time versus non-constant quantitiy over a continuous unit of time:

a. *Olen ostanut jo kauan sähkö-n tältä yhtiöltä.*
AUX.1SG buy.PSTAP already long electricity-GEN this.ABL company.ABL
‘I have been buying all my electricity from this company already for a long time.’

b. *Olen ostanut jo kauan sähkö-ä tältä yhtiöltä.*
AUX.1SG buy.PSTAP already long electricity-PAR this.ABL company.ABL
‘I have been buying some of my electricity from this company already for a long time (but I have been buying also from other companies).’ (Huumo 2006: 531)

(46) Pragmatic contrast:

a. *Lainaa nenäliina-si!*
borrow handkerchief-POSS.2SG
‘Lend (me) your handkerchief!’

b. *Lainaa nenäliina-a-si!*
borrow-IMP handkerchief-PAR-POSS.2SG

'Lend (me) your handkerchief (and I will give it back to you in a moment)!' (Kangasmaa-Minn 1999: 48)

It seems that DOM is governed by two main functions. The first is quantification of the object – the Total object is assigned a definite or constant quantity over time, and the Partitive object is assigned an indefinite amount of the object.

In instances in which measuring the object does not make sense (because the amount of object over time must be constant due to the meaning of verb and the semantic type of object), object marking may be involved in theme-to-event homomorphism.

This phenomenon has been examined in some studies. Kiparsky (2005) who elaborates on object marking in Finnish, calls it, *gradability*, Tamm (2012) who deals with Estonian verbs, refers to it as *scalarity*. If a situation in a sentence reaches "the perfect state" (Huumo 2006) the Total object is used, while "the imperfect state" is marked with the Partitive. The Partitive object, thus, applies to situations which are not characterised by the "perfect state". Such perfect states are represented in (42a) and (41a).

Reaching the perfect state in a situation is similar to what Lindstedt (1995: 97) defines as MATERIAL BOUND: "something which forms the natural end-point of a situation by virtue of its being just the kind of situation it is."[21]

Lindstedt (1995: 97) also identifies a second, TEMPORAL type of bound, where the end point of the situation is defined temporally: hence, it is possible for situations which do not have a material bound.[22] Such temporal bounds lie behind the semantics of Romance aspect and can be expressed lexically with temporal expressions of durative temporalisation.

As shown in Section 4.5.3, expressing durative temporalisation with Total object clauses is more restricted than with a Partitive object. Thus, purely temporal bounding is not enough to use a Total object in a sentence.

When neither object quantification nor situation quantification is relevant in the sentence, DOM takes over other functions of the pragmatic as in (46).

4.3.4 Resultative constructions

RESULTATIVE CONSTRUCTIONS (Hakulinen et al. 2004: §482) are three-element structures requiring the subject, object and an adverbial (usually in the Translative

[21] Material bound is proximate to Garey's 1957 *telicity* – a potential, natural end point.

[22] Following the terminology of Garey (1957), they are atelic.

or Illative) or a free ad-verbal. They are common in both transitive (47a) and intransitive verbs (47b). Often the type of object differs somewhat from the normal usage as in (47c), where normally the name of a language or the content of the talk would be expected:

(47) a. *Mies hakkasi jää-n puhki.*
man pound.SPST ice-GEN through
'The man pounded the ice through' (Leino 2010: 118)

b. *Kävelin jalka-ni puhki.*
walk.SPST.1SG feet-POSS.1SG through
'I walked my legs down'

c. *Puhuin kurkku-ni käheäksi.*
talk.SPST.1SG throat-POSS.1SG hoarse-TRANS
'I talked my throat hoarse' (Leino 2010: 120)

Thanks to the adverbial it is possible to discretise the TSIT, as the situations above may last only until 'legs are down' or the 'throat is talked hoarse'. The Total object is possible in resultative constructions, even with verbs which normally take only the Partitive object, which would confirm that indeed 'the default' usage of the Partitive is related to measuring out the situation as 'imperfect':

(48) a. *Kalle halasi Liisa-a.*
Kalle hug.SPST Liisa-PAR
'Kalle hugged Liisa.'

b. *Kalle halasi Liisa-a rauhalliseksi.*
Kalle hug.SPST Liisa-PAR calm-TRANS
'Kalle was hugging Liisa so she would became calm.' (Leino 2010: 121)

c. *Kalle halasi Liisa-n rauhalliseksi.*
Kalle hug.SPST Liisa-GEN calm-TRANS
'Kalle hugged Liisa so she became calm.'

4.3.5 Measure adverbials in the object cases

Measure adverbials in the object cases (Fin. *objektinsijaiset määrän adverbiaalit*, abbreviated henceforth following the Finnish linguistic tradition to OSMA, Tuomikoski 1978) are adverbials which morphologically resemble the object, but are also measures of size, degree, lasting in time, or distance.

(49) a. *Luin kirja-a tunni-n.*
read.SPST.1SG book-PAR hour-GEN
'I read the/a book for/in an hour.'

b. *Juoksin kaksi kilometri-ä.*
run.SPST.1SG two kilometre-PAR
'I ran two kilometres.'

c. *Dieselin hinta nousee vuoden alussa jopa 12 sentti-ä litralta.*
diesel.GEN price increase.3SG year.GEN beginning.INE even 12 cent-PAR litre.ABL
'In the beginning of the year, the price of diesel will increase even 12 cents per litre.'[23]

Nevertheless, OSMAs differ syntactically in comparison to the object, for example (Vilkuna 2000: 85) writes that they cannot belong to the agentive participle phrase:

(50) a. *Piirtelin ukonkuvia koko tunnin.*
draw.SPST.1SG figure.picture.PL.PAR whole hour-GEN
'I was drawing pictures of figures for the whole hour.'

b. *Piirtele-mä-ni ukonkuva-t olivat elämäni onnistuneimpia.*
draw-AG-poss.1SG figure.picture-PL.NOM COP.SPST.3PL life.POSS.1SG most.successful.PL.PAR
'The figures I drew were among the best in my life.'

[23] https://yle.fi/uutiset/3-5475332

c. * *Piirtele-mä-ni tunti oli elämäni*
draw-AG-POSS.1SG hour COP.SPST life.POSS.1SG
ikävästyttävimpiä.
most.tedious.PL.PAR
('The hour I was drawing was one of the most tedious in my life.')

Unlike objects, several OSMAs may appear in a sentence:

(51) *Hän juoksi* ***viisi kilometri-ä*** *joka päivä* ***koko kesä-n***.
(s)he run.SPST five kilometre-PAR every day whole summer-GEN
'During the whole summer, (s)he ran five kilometres every day.' (Hakulinen et al. 2004: §972)

Usually OSMAs do not belong to the predicate-argument structure. Some verbs, however, require an obligatory argument of measurement:

(52) a. *Illallinen maksoi 20 euro-a.*
dinner cost.SPST 20 euro-PAR
'The dinner cost 20 euros.'

b. *Kokous kesti tunni-n.*
meeting last.SPST hour-GEN
'The meeting lasted an hour.'

OSMA is an important mean of durative temporalisation in Finnish, discussed further in Section 4.5.

4.3.6 Spatial expressions

The temporal progress of a situation can be measured on a path scale (see Section 2.6.4). The path scale requires a moving argument which appears in the sentence as the subject or the object, as well as an argument which defines the path. The properties of the path can be defined by naming the path or by marking some landmarks belonging to the path or its border points.

Finnish has a very diverse system of marking spatial relations with cases (Section 4.2.2) and with adpositions.

(53) a. *Juoksin koti-in.*
run.SPST.1SG home-ILL
'I ran/was running home.'

b. *Juoksin koti-in asti.*
run.SPST.1SG home-ILL until
'I ran all the way home.'

(54) a. *Juoksin kohti koti-a.*
run.SPST.1SG towards home-PAR
'I was running home.'

b. *Juoksin koti-in päin.*
run.SPST.1SG home-ILL towards
'I was running home.'

TSIT can be homomorphic with the path, but it does not necessarily have the same scalar properties as the path-scale argument. Thus, in (53b) it is assignable to discrete TR, because the length of the path represents a discrete value, while in (54) it can take any value between 0 and the length of the path leading home. (53a) seems neutral as the interpretation depends on the context, as in the examples below:

(55) a. *Kun juoksin koti-in, kompastuin ja kaaduin.*
when run.SPST.1SG home-ILL stuble.SPST.1SG and fall.SPST.1SG
'When I was running home, I stumbled and fell down.'

b. *Juoksin koti-in, hain avaimen ja juoksin*
run.SPST.1SG home-ILL take.SPST.1SG key.GEN and run.SPST.1SG
takaisin kouluun.
back school.ILL
'I ran home, took the key, and ran back to school.'

However, as shown in Section 4.4.1, the way path-scale argument is expressed constrains the type of temporal and lexical localising expression.

When the mover is expressed by the object, the assignability of TSIT to the type of scale is marked with DOM:

(56) a. *Mies siirsi pöydä-n keittiöön.*
man move.SPST table-GEN kitchen-ILL
'The man moved the table to the kitchen.'

b. *Mies siirsi pöytä-ä keittiöön.*
man move.SPST table-PAR kitchen-ILL
'The man was moving the table to the kitchen.'

The morphological derivation, which in Finnish reduces or extends the number of arguments, influences the possibility of marking theme-to-event homomorphism with DOM as it is in (56) in contrast to (57):[24]

(57) a. *Pöytä siirt-y-i keittiö-ön.*
table move-REFL-SPST kitchen-ILL.
'The table moved to the kitchen.'

b. * *Pöytää siirt-y-i keittiö-ön.*
table.PAR move-REFL-SPST kitchen-ILL.
('The table was moving to the kitchen.')

However, in the case of an uncountable subject such differentiating is possible (although not entirely unambiguous, see Section 4.6.2).

(58) a. *Raha siirt-y-i ulkomai-lle.*
money move-REFL-SPST abroad-ALL.
'The money moved abroad.'

b. *Raha-a siirt-y-i ulkomail-lle.*
money.PAR move-REFL-SPST abroad-ALL.
'Some money moved/was moving abroad.' or 'The money was moving abroad.'

[24]The additional interference between number and countability on theme-to-event homomorphism is discussed in Section 4.6.2.

4.4 Temporal localising

4.4.1 Deixis

Tenses in Finnish and the temporal reference

In Finnish, four formal tenses can be distinguished: two simple and two analytical. The Non-past tense is unmarked, while the Simple Past is marked with the morpheme *-i-* or the allomorph *-si-* in affirmative sentences, and with past participles in negated sentences.

The analytical tenses in Finnish are the Perfect and the Pluperfect. Both consist of the auxiliary formed from the verb *olla* ('to be') and the past participle. The auxiliary conjugates in the Pluperfect according to the Simple Past paradigm and in the Perfect according to the Non-past tense paradigm.

The Simple Past and the Pluperfect localise situations in the past. The Simple Past uses an absolute time span, the Pluperfect underlines an order of situations, so it requires a minimal context of another sentence clause or some background context and it is also a taxis marker (see Section 4.4.2):

(59) *Lu-i-n* *lehden.*
read-SPST-1SG newspaper.GEN
'I read the newspaper.'

(60) (Välillä Antin oli vaikea kestää työhöni liittyvää vaaraa.)
Olimme *tutustuneet* *murhan* *merkeissä.*
AUX-SPST-1PL become.familar-PSTAP.PL murder.GEN sign.PL.INE
'(From time to time Antti found it difficult to stand the danger related to my profession.) We had met in the murder circumstances.' (S1515)

The Non-Past may refer to both the present moment and the future. Distinguishing between these two types of temporal reference is possible due to temporal adverbials, the case of the direct object or other elements, in particular pragmatic context:

(61) a. *Luen* *lehde-n.*
read.1SG newspaper-GEN.
'I will read the newspaper.'

b. *Luen lehte-ä.*
read.1SG newspaper-PAR.
'I am reading the/a newspaper.' or 'I read a newspaper'

(62) a. *Menen lääkäriin.*
go.1SG doctor.ILL
'I am going to the doctor.'
b. *Huomenna menen lääkäriin.*
tomorrow go.1SG doctor.ILL
'Tomorrow, I will go to the doctor.'

Also the periphrastic construction *tulla* + V-INF3-ILL[25] appears mostly in the written language; it is used for forecasting future events:

(63) *Kaisa tule-e asu-ma-an täällä kolme vuotta.*
Kaisa AUX-3SG live-INF3-ILL here three year.PAR
'Kaisa is going to live here for three years.'

The Perfect fulfils all the functions described by Comrie (1976: 56–61)

(64) a. Perfect of result (the result of the situation is relevant at TU):
Olen siivon-nut huonee-ni.
AUX.1SG clean-PSTAP room.POSS.1SG
'I have cleaned my room.' (and it is clean now)

b. Perfect of recent past (according to the speaker's perception, the situation has taken place shortly before the TU) :
Hän on jo syö-nyt.
he AUX.1SG already eat-PSTAP
'He has already eaten.'

c. Experiential ("a given situation has held at least once during some time in the past leading up to the present" Comrie 1976: 58):

[25]Finnish has a complex system of infinitives, where the third infinitive can appear in spatial cases and the Abessive.

Olen käy-nyt Pariisissa.
AUX.1SG walk-PSTAP Paris.INE

'I have been to Paris.'

d. Persistent situations that are valid also at TU:

Olen asu-nut täällä viisi vuotta.
AUX.1SG live.PSTAP here five year.PAR

'I have lived here for five years.'

e. Future perfect:

Huomenna tähän aikaan olen jo kirjoitta-nut tenttini.
tomorrow this.ILL time.ILL AUX.1SG already write-PSTAP exam.POSS1SG

'Tomorrow at that time, I will have sat my exam.'

Lexical temporal localising

Most mechanisms describing lexical temporal localising in Polish hold also in Finnish. Names of events, calendar dates etc. assign a TSIT to a TR on the time axis. The TR to which the TSIT is assigned can be pointed out independently (67) or at a distance from some other point (66). In particular it may be identical with TU (65):

(65) *Aloitamme kokouksen* ***nyt***.
begin.1PL meeting.GEN now
'We begin the meeting now.'

(66) *Aloitamme kokouksen* ***tunnin kuluttua***.
begin.1PL meeting.GEN hour pass.PSTPP.PAR.
'We beginning the meeting in an hour.'

(67) *Aloitamme kokouksen* ***kello yhdeksän***.
begin.1PL meeting.GEN clock nine

‘We begin the meeting at nine o’clock.

Temporal expressions participating in temporal localising have three forms in Finnish:

- bare adverb (e.g. *eilen* ‘yesterday’)
- declined noun (e.g. *tammikuu-ssa* ‘January-INE, in January’)
- adverbial noun phrase (e.g. *viikon aika-na* ‘week.GEN time-ESS, during the week’)
- temporal particle (e.g. *jo* ‘already’)

Most cases of spatial location are in use for marking temporal localising:

- the Inessive (e.g. *tammikuu-ssa* ‘in January’)
- the Essive (e.g. *vuon-na 2016* ‘in the year 2016’)
- the Adessive (e.g. *kesä-llä* ‘in the summer’)
- the Illative (e.g. *päiväsaika-an* ‘in the daytime’)
- the Translative (e.g. *kuude-ksi* ‘for nine o’clock’)
- the Elative (e.g. *alkuilla-sta* ‘in the early evening’)
- the Ablative (e.g. *yhdeksä-ltä* ‘at nine o’clock’)

4.4.2 Taxis

Tense as taxis marker

Taxis in Finnish has received very little research attention. It has been only recently described by Tommola (2016) though even this work does not reflect all problems. Similarly to Polish, grammatical tenses are the main means of expressing simultaneity, anteriority and posteriority. In general, when two clauses appear in the same tense, they are interpreted as simultaneous. The Perfect and the Pluperfect mark anteriority.

(68) a. simultaneity

Anna nukku-u, kun Ella keittä-ä puuroa.
Anna sleep-3SG while Ella cook-3SG porridge.PAR
'Anna is sleeping while Ella is cooking porridge.'

b. anterior

Anna herä-ä, kun Ella on keittä-nyt puuron.
Anna wake.up-3SG when Ella AUX cook-PSTAP porridge.GEN
'Anna will wake up, when Ella has cooked the porridge.'

c. simultaneity

Anna nukku-i, kun Ella keitt-i puuroa.
Anna sleep-SPST while Ella cook-SPST porridge.PAR
'Anna was sleeping while Ella was cooking porridge.'

d. anterior

Anna heräsi, kun Ella ol-i keittä-nyt puuron.
Anna wake.up-SPST when Ella AUX-SPST eat-PSTAP porridge.GEN
'Anna woke up when Ella had cooked the porridge.'

A Non-past conditional form may be a marker of the posterior, when it is used as relative to a situation of a past temporal reference, but the use of conditional has modal colouring:

(69) *Katia toivo-i, että paikalla ol-isi vain Matias.*
Katia hope-SPST that place.ADE be-COND only Matias
'Katia hoped that only Matias would be there.' (Ilves 2016: 20)

Non-finite taxis markers

In addition to tenses, some non-finite forms are specialised in marking taxis:

(70) a. Present Participle (simultaneity of clauses)

Kuulin sinun ***laula-va-n***
hear you.GEN sing-PP-GEN

'I heard how you were singing.'

b. Active Past Participle (anterior or partial overlap of the TR of TSIT of the non-finite clause and TR of TSIT of the finite clause)

Kuulusteluissa paljastui Hitlerin ilmeisesti
interrogation.PL.INE reveal.REFL.SPST Hitler.GEN obviously
sairastaneen Parkinsonin tautia.
be.ill-PSTAP-GEN Parkinson.GEN disease.PAR

'During the interrogations it became clear that Hitler had suffered from Parkinson's disease.' (Hakulinen et al. 2004: §538)

c. The second infinitive Inessive + Possesive (TR of TSIT of the finite clause included in the TR of TSIT expressed in the non-finite clause)

Olsson löysi Saariahon elokuvamusiikin
Olsson find.SPST Saariaho.GEN soundtrack.GEN
siivo-te-ssa-an *varastoaan.*
clean-INF2-INE-POSS3 storehouse.PAR.POSS3

'Olsson found Saariaho's soundtrack when he was cleaning his storehouse.'[26]

d. Passive Past Participle + Possesive (TR of TSIT of non-finite clause anterior to TR of TSIT of finite clause)

Laule-ttu-a-ni *menin kotiin.*
sing-PSTPP-PAR-POSS.1SG go.SPST.1SG home.ILL

'Having sung, I went home.'

e. the second infinitive instructive (simultaneity of clauses)

[26] http://www.hs.fi/paivanlehti/12042017/art-2000005167033.html

f. *Nyt Moseley kutoo pipon puole<ssa>toista tunnissa*
now Moseley knit.3SG beanie.GEN one.and.a.half<INE> hour.INE
televisio-ta ***katsell-e-n***.
tv.PAR watch-INF2-INSTR
'Nowadays, Moseley knits a beanie in one and a half hours while watching the TV.' [27]

Taxis and lexical expressions

Taxis in Finnish is also expressed lexically, and, as in Polish, finer distinctions are made with the usage of appropriate temporal conjunctions (usually containing the expression *kun* 'when') (71) and adverbials (72). Contact posteriority (posterior TR immediately after the anterior one) is expressed by conjunctions combined with an adverb *heti kun* and *kohta kun* ('as soon as') as in (73).

(71) a. *Tulen,* ***kun*** *olen syönyt.*
come.1SG when AUX.1SG eat.PSTAP
'I will come, when I have eaten.'

b. ***Kun*** *lahjoitus oli tilillä, yhdistyksen toiminta*
when donation be.SPST account.ADE, association.GEN activity
saatiin käyntiin
get.PASS.SPST run.ILL
'When the donation was on the account, the activity of the association started running.'[28]

(72) a. *Ensin tuijotimme rusinaa kymmenen minuuttia,* ***sitten***
first stare.SPST.1PL raisin.PAR ten minute.PAR then
haistelimme sitä yhtä kauan.
smell.SPST.1PL it.PAR as long
First, we stared at a raisin for ten minutes, then we smelt it for an equally long time' (Ilves 2016: 70)

[27] http://www.iltamakasiini.fi/artikkeli/459297-86-vuotias-mies-kutoo-pipoja-keskosille

[28] http://www.kielitoimistonohjepankki.fi/haku/kun/luokka/361/ohje/92

b. *Avaan jo suuni väittääkseni vastaan, mutta*
open.1SG already mouth.POSS.1SG claim against, but
juuri silloin *Tainan pöytäpuhelin soi.*
exactly then Taina.GEN deskphone ring
'I am already opening my mouth to object, but exactly then Taina's desk phone is ringing.' (Ilves 2016: 209)

(73) a. *Tilaisuus tuli yllättäen,* ***juuri kun*** *hän oli*
opportunity come.SPST surprisingly exactly when (s)he AUX.SPST
luullut kaiken olevan menetetty.
think.PSTAP everything.GEN AUX.PP.GEN loose.PSTPP
'An opportunity appeared surprisingly, exactly when she had thought everything had been lost.' (Ilves 2016: 268)

b. *Pohjoiskarjalainen hankkii ajokortin* ***heti kun***
North.karelian get.3SG driving.licence.GEN immediately when
pystyy
able.3SG
'A North Carelian gets a driving licence as soon as he can.'[29]

4.5 Durative temporalisation

4.5.1 The Finnish progressive

The combination *olla* + V-INF3-INE is sometimes compared to the progressive function of the English Present Participle (Vilkuna 2000: 256). Besides the progressive meaning (74a), Vilkuna (2000: 257) distinguishes temporary presence in some place for some purpose (74b) and the nearly terminating situation, sometimes called *proximative* (74c):

(74) a. *Olen kirjoitta-ma-ssa artikkelia.*
AUX.1SG write-INF3-INE article.PAR
'I am writing an article.'

[29] http://yle.fi/uutiset/3-6715024

b. *Pyykki on narulla kuivu-ma-ssa.*
laundry AUX rope.ADE dry-INF3-INE
'The laundry is drying on the rope.'

c. *Hanna oli voitta-ma-ssa juoksun.*
Hanna AUX.SPST win-INF3-INE run.GEN
'Hanna was about to win the run.'

Similarly to the progressive in English which cannot be used with all verbs, INF3.INE is rarely used with such verbs as *rakastaa* 'to love', or *sijaita* 'to be located, situated'. Interestingly, the lexicalised form *olla olemassa* (from *olla*, 'to be') is possible and it means 'to exist'.

In the case of the verbs of movement *mennä* ('to go'), *lähteä* ('to leave') and *tulla* ('to come') the lexicalised noun constructions *olla menossa*, *olla lähdössä* and *olla tulossa* are more frequently in use than the third infinitive.

Nevertheless, the usage of progressive forms is not obligatory in the same contexts as in English as for example in taxical relations:

(75) a. *Lu-i-n kirjaa, kun puhelin soi.*
read-SPST-1SG book.PAR, when telephone ring.SPST
b. *Ol-i-n luke-ma-ssa kirjaa, kun puhelin soi.*
AUX-SPST-1SG read-INF3-INE book.PAR when telephone ring.SPST
'I was reading a/the book when the telephone rang.'

4.5.2 Lexical expressions

Durative temporalisation is realised lexically in:

- phrases naming the interval, usually in the form of OSMA, such as *kaksi tuntia* 'two hour.PAR', *viikon* 'week.GEN'
- specifying lasting qualitatively with OSMA, such as *kauan* 'for a long time' or in the Illative, such as *pitkään* 'long'

A similar effect to Polish durative temporalisation can be achieved by specifying the boundaries of lasting. The boundaries are expressed either in the Elative and Illative forms of bare nouns or adpositional phrases with the particles *saakka* and *asti*, which are near synonyms 'since; until; as far as', or *lähtien* 'since'. The

difference between bare noun and adpositional phrase is rather fine as described by Päiviö (2007).

(76) a. *Olen Suomessa maanantai-hin asti.*
be.1SG Finland.INE Monday-ILL as.long.as
'I am in Finland until Monday.'

b. *Olen Suomessa maantai-hin.*
be.1SG Finland.INE Monday-ILL
'I am in Finland until Monday.'

a. *Ei ole enää mitään tekosyytä tehdä töitä*
NEG be yet something.PAR.NEG excuse.PAR do.INF1 job.PL.PAR
aamu-sta ilta-an.
morning-ELA evening-ILL
'There is no excuse to work from dusk till dawn.'[30]

b. *Yle lähettää suorana aamu-sta ilta-an asti.*
Yle send.3SG straight.ESS morning-ELA evening-ILL as.far.as
'Yle is broadcasting live from dusk till dawn.'[31]

4.5.3 Co-occurrence restrictions between durative adverbials and the theme-to-event homomorphism

OSMA-type durative adverbials can only occur in clauses where the theme-to-event homomorphism allows the assignment of TSIT to continuous TR.

(77) a. *Luin kirja-a kaksi päivä-ä.*
read.SPST.1SG book-PAR two day-PAR
'I read a/the book for two days.'

b. * *Luin kirja-n kaksi päivä-ä.*
read.SPST.1SG book-GEN two day-PAR

[30] https://www.firstbeat.com/fi/menestystarinat/ei-ole-enaa-mitaan-tekosyyta-tehda-toita-aamusta-iltaan/

[31] http://yle.fi/uutiset/osasto/sapmi/saamenmaa_juhlii__yle_lahettaa_suorana_aamusta_iltaan_asti/9444068

('I read the book in two days.')

(78) a. *Juoksin tunni-n.*
run.SPST.1SG hour-GEN
'I ran for an hour.'

b. ? *Juoksin koti-in tunni-n.*
run.SPST.1SG home-ILL hour-GEN
('I ran home in an hour')

On the other hand, Finnish temporal adverbials in the Inessive combine with the Total object, but not with the Partitive, which closely resembles the strong preference for PFV when the Polish *w*+ACC is used.

(79) a. *Luin kirja-n reilu-ssa pari-ssa päivä-ssä (ja sitä*
read.SPST.1SG book-GEN proper-INE pair-INE day-PL-INE
juuri tarvitsin)

'I read the book in a good two days (and I needed exactly that).'[32]

b. *Nyt Moseley kutoo pipo-n puole<ssa>toista tunni-ssa*
now Moseley knit.3SG beanie-GEN one.and.a.half<INE> hour-INE
televisiota katsellen.
tv.PAR watch.INF2.INSTR
'Nowadays, Moseley knits a beanie in one and a half hours while watching the TV.' [33]

c. * *Luin kirja-a kahde-ssa päivä-ssä.*
read.SPST.1SG book-PAR two-INE day-INE
('I read a/the book for two days.')

Also clauses containing a path-scale argument allow the Inessive.

[32] http://uino.indiedays.com/2016/10/20/nollaamista/
[33] http://www.iltamakasiini.fi/artikkeli/459297-86-vuotias-mies-kutoo-pipoja-keskosille

(80) *Juoksin koti-in tunni-ssa.*
run.SPST.1SG home-ILL hour-INE
'I ran home in (within) an hour.'

It appears that the constraints of usage on the Illative are even stronger than those for the Inessive:

(81) a. * *Juoksin koti-in tunti-in.*
run.SPST.1SG home-ILL hour-ILL
('I ran home in (within) an hour.')

b. *Juoksin koti-in asti tuntiin.*
run.SPST.1SG home-ILL until hour-ILL
'I ran (all the way) home in (within) an hour.'

(82) a. * *Luin kirja-a päivä-än.*
read.SPST.1SG book-PAR day-ILL

b. *Luin kirja-n päivä-än.*
read.SPST.1SG book-GEN day-ILL
'I read the book in a day.'

Although the limitations on transitive constructions are similar to those applying to the Inessive, the Illative cannot be used together with the path-scale argument, unless reaching the end point of the path is explicitly stated.

4.6 Pluractionality

4.6.1 Temporal quantification of Finnish sentence

Similarly to Polish, Finnish does not have any separate verbal marking of existential and universal quantification.

Nonetheless, the Perfect often expresses non-specific quantification, as the focus of the localising of TSIT in the Perfect function of result, experiential or recent past (see above, Section 4.4) is related to some moment before the TU, while the specific location of the TR to which TSIT is assigned is out of focus.

Universally quantified sentences are mostly formulated in the Non-past. In that context, sentences often take a subjectless third person singular form, the *missing person* (Hakulinen & Karttunen 1973) as in (83a), or impersonal (83b):

(83) a. *Kotona ikävästyy.*
home.ESS be.bored.REFL.3SG
'One gets bored at home.'

b. *Englannissa ajetaan tien vasentaa laitaa.*
England.INE drive.PASS road.GEN left.PAR side.PAR
'In England one drives on the left side of the road.'

The missing-person sentence, is quite restricted as to its usage in comparison to impersonal constructions:

(84) * *Englannissa ajaa tien vasentaa laitaa.*
England.INE drive road.GEN left.PAR side.PAR
('In England one drives on the left side of the road')

This constraint can be partly explained by politeness strategies, which, according to Hakulinen (1987), in Finnish include hiding the agent. The missing-person construction has conventionalised to the function of generalisations applying to the speaker or the addressee of the utterance.

The agent can be hidden in universal quantification by detransitivisers as in example (3), discussed in Section 4.2.1:

(3) *Automatisoitu ovi avautuu helposti nappia painamalla*
automatic door open.REFL.3SG easily button.PAR press.INF3.ADE
(...) .

'Automatic doors open easily when pressing the button.'[34]

4.6.2 Quantification over referents

Finnish does not have a category of quantifying classifiers similar to articles. Vilkuna (1992) has written probably the most exhaustive monograph on definite-

[34] http://www.jkovet.fi/talvikamppanja-tasmalaake-pakkaseen/

ness and referentiality in Finnish. She concludes that those categories are relative, as they require always accounting for the pragmatic or discourse-contextual knowledge of the speaker and the hearer.

However, the choice of cases for subject and object marking enables some differentiation in quantification in the case of uncountable referents as in (38) from Section 4.3.3 where the plural usage implied an unknown number of letters, or in singular in the example (58b) discussed in Section 4.3.6:

(38a) *Tämä laatikko sisältää vanha-t kirjee-ni.*
this box contain.3SG old-PL.NOM letters-POSS.1SG
'This box contains all of my old letters.'

(38b) *Tämä laatikko sisältää vanho-j-a kirje-i-tä-ni.*
this box contain.3SG old-PL-PAR letter-PL-PAR-POSS.1SG
'This box contains some of my old letters.'

(58b) *Raha-a siirt-y-i ulkomail-lle.*
money-PAR move-REFL-SPST abroad-ALL
'Some money moved/was moving abroad.' or 'The money was moving abroad.'

In the latter example, two-fold interpretation is possible. Either *some (amount of) money was transferred abroad*, or *the money was being transferred*. This problem applies also to sentences were the plural Partitive subjects and objects are used indicating either that the material bound is not reached in the situation, or the quantity of subject or object is unknown, but the material bound is reached.

Following Kangasmaa-Minn (1978: 19, 27) one can compare the following pairs of sentences:

(85) a. *Ihmiset kuolivat nälkään.*
human.PL.NOM die.SPST.3PL hunger.ILL
'People died of hunger.'[35]

b. *Ihmisiä kuoli nälkään.*
human.PL.PAR die.SPST.3SG hunger.ILL

[35]This example is not unambiguous neither, and I return to it in the next subsection.

'(Some) people died/ people were dying of hunger.'

(86) a. *Isä kantoi lapset kotiin.*
father take.SPST child.PL.NOM home.ILL
'The father took the children home.'

b. *Isä kantoi lapsia kotiin.*
father take.SPST child.PL.PAR home.ILL
'The father took some children/ was taking the children home.'

In these cases, the Partitive plural form of subject or object is the source of ambiguity between the theme-to-event homomorphism and the unspecified quantity of an argument. These kinds of uncertainty are usually narrowed in Finnish through the context.

4.6.3 Specifying type of frequency

Type of frequency can be overtly expressed in Finnish only with adverbial pharases and involves various morphological forms:

- summaric (e.g. *kerran* 'time.GEN, once', *kolmasti* 'three times')
- specific cycle type (e.g. *kerran päivässä* 'once in a day', joka päivä 'everyday', *päivittäin* 'daily', *päivisin* '(always) in the daytime')
- unspecific cycle type (e.g. *yleensä* 'usually', *joskus* 'sometimes', *harvoin* 'rarely', *välillä* 'from time to time')

Some expressions, such as *kerran*, belong to OSMA:

(87) a. *Stomil nipisti minua,* ***kaksi kertaa*** *vielä.*
Stomil pinch.SPST me.PAR two time.PAR still
'Stomil pinched me, twice more.' (S1720)

b. *joka on sittemmin laajennettu jo* ***useita kertoja***
which AUX afterwards extend.PSTPP already several.PL.PAR time.PL.PAR

'which has been being extended already several times.' (S1033)

Although some expressions seem synonymous, they are constrained as to their usage:

(88) a. *Sunnutaisin harjaa hampaat.*
on.Sundays brush teeth.PL.NOM
'On Sundays one brushes one's teeth.'

b. * *Joka sunnutai harjaa hampaat.*
every Sunday brush teeth.PL.NOM
('Every Sunday one brushes one's teeth.') (Hakulinen & Karttunen 1973)

The difference between *sunnuntaisin* and *joka sunnuntai* is according to Hakulinen & Karttunen (1973) analogical to the difference between *any* and *every* that causes the usage restrictions of *joka sunnuntai* in the generic context.

Type of frequency is not expressed by verbal forms, and as shown by Lindstedt (1984: 24), no association between DOM and the type of frequency can be observed:

(89) a. Kesällä luin joka päivä venäläise-n kirja-n.
summer.ADE read.SPST.1SG every day Russian-GEN book-GEN

'In the summer I read through a/the Russian book every day.'

b. Kesällä luin joka päivä venäläis-tä kirja-a.
summer.ADE read.SPST.1SG every day Russian-PAR book-PAR

'In the summer I read/was reading some Russian book every day.'

DOM in the examples above concerns the quantity of the book(s) read. Thus, in (89a) everyday one book is read, while in (89b) only some portion of some book is read.

Let me now return to the examples (85a, 86a), which are ambiguous with reference to the type of frequency. In those examples the plural form of the subject

or object does not allow for stating how the situation of dying or carrying took place. It is not clear whether people or children participated in the situation one after another, or whether they all died/were carried once. Thus, the Total object and Nominative subject are the result of non-gradability in situation, but the inner frequency structure cannot be distinguished.

4.7 Conclusions

As shown above, many mechanisms of expressing temporality in Finnish indicative clauses are similar to Polish. Both languages have a very reduced system of expressing non-past temporal reference, and benefit from using lexical expressions of temporal localising and durative temporalisation. Neither Finnish nor Polish have a system of articles which could help in interpreting the pluractionality of the sentence.

Nevertheless, Polish has verbal aspect expressed with prefixes and suffixes which interact in each domain of temporality, for example, as future reference PFV, IPFV of simultaneity or habitual.

Finnish, in contrast, has more developed system of past tenses and Paritive-Total object opposition, which to some extent resembles the aspectual mechanism of Polish, since Total object rather restricts the properties of TSIT. Nevertheless, since DOM operates on the nominal and not on the verbal phrase, ambiguities concerning type of quantification (discrete measurement of the nominal object or of TSIT) are hard to avoid.

Apart from DOM, the developed system of spatial cases used for marking oblique arguments seems to play some role, in particular in resultative constructions, so in marking the explicit material bound.

In order to examine the interactions between different levels of the Finnish and Polish systems with PVA as the key feature in mind, I turn now to the quantitative analysis of written, parallel texts.

Chapter 5

The Polish-Finnish parallel corpus

5.1 The corpus-oriented approach

All statements made here concerning the functional correlates of PVA in Finnish are drawn on the basis of empirical material collected in a parallel corpus.

It is important to distinguish two meanings of the term *corpus* in contemporary linguistics. The first, broader sense, CORPUS_1 is used with reference to text collections of a defined size and content, often considered representative for some type of language variety, like national corpora or reference corpora. In a narrow sense, a CORPUS_2 is a collection of texts representative with regards to a research hypothesis or research questions.

The main advantage of the corpus-oriented approach, in comparison with traditional studies based on researchers' intuition, is the higher level of *external validity* – the degree to which the results of the study can be generalised and replicated. Statements based on the collection of a wide range of utterances produced in the natural environment by different speakers are less likely to be subject to bias.

Secondly, Johansson (2007: 51) argues that "through corpora we can observe patterns in language which we were unaware of before or only vaguely glimpsed. (...) this applies particularly to multilingual corpora. We can see more clearly what individual languages are alike, what they share and – perhaps eventually – what characterises language in general."

PARALLEL CORPORA, where the original text is stored together with its translations, can be particularly useful for contrastive and cross-linguistic studies. For example Xiao & McEnery (2006) examine how temporal and aspectual meanings are expressed in English–Chinese translations, and Santos (1996) uses parallel

corpora for mapping temporal categories between Portuguese and English.

In linguistic typology, as advocated for example by Dahl (2007) and Wälchli (2007), parallel corpora have become a popular and fruitful type of data.

Mihalcea & Simard (2005: 239) suggest that the translation of a text in another language can be seen as a semantic representation of that text so new methods of studying language can be applied.

In this chapter, I describe the most general qualitative and quantitative features of the corpus_1 collected for the purpose of the current study, and corpus_2 – the sample of corpus_1 containing exclusively clauses relevant to the subject of the present study. I discuss the matters of representativeness and balance as well as external factors which influenced the structure of the corpus.

5.2 Preliminary sampling criteria

5.2.1 Age and language of texts

The corpus compiled for this study is a parallel corpus_1, a text collection consisting of original texts in Finnish and Polish and their Polish and Finnish translations. All texts were collected in 2013 and 2014. In order to appear in the corpus, texts must have been written after the Second World War, be direct translations, and the source text must have been written originally in one of the languages in focus. All texts and their translations occurred naturally – for real-life purposes, and to my best knowledge they were produced by professional translators.

5.2.2 Smoothing the effect of interference

Numerous studies (e.g. Baker 1999; Chesterman 2004, 2010; Kenny 1998; Toury 1995) discuss the matter of TRANSLATIONAL UNIVERSALS, that is, features that indicate the text is a translated one, such as changing the length of the text, the law of interference, standardisation, simplification or conventionalisation.

From the point of this study, the law of interference ("phenomena pertaining to the make-up of the source text tend to be transferred to the target text" Toury 1995: 275) is particularily interesting. To date some studies which empirically confirm the law of interference, McEnery et al. (2006: 321 – 343) show that the frequency of aspectual markers in Mandarin Chinese is significantly lower in the texts translated from English than in the original Mandarin Chinese texts. Eskola (2004) concludes that the frequencies of non-finite temporal constructions

which do not have straightforward counterparts in other languages are higher in original Finnish texts than in translations from English and Russian. Similar conclusions are drawn in Tirkkonen-Condit (2004) as to unique for Finnish verbs of sufficiency and the clitic pragmatic particles =*kin* and =*hAn*. Additionally, Eskola (2004) observes that Finnish final constructions, which are common in other languages, have higher frequency in Finnish translations than in original texts. Both authors propose that target language-specific elements which do not have equivalents in the source language tend to be under-represented in translation, whereas constructions with more straightforward counterparts are over-represented, when the source language does not use competitive forms.

Based on the sketches of temporal systems from Chapters 3 and 4, there are premises to assume that the law of interference applies to the temporal domain in the context of translations between Finnish and Polish, as the grammatical systems differ a lot. In order to avoid that it affects the quantitative results in this study, I follow the recommendation of Ebeling (1998). The corpus is bidirectional, that is, both Finnish and Polish contribute equally as source languages. Balancing the corpus in terms of directionality of translation should smooth the potential interference to some extent.[1]

It is essential to underline that the current study deals with functional linguistic equivalents and not with translational equivalents. Thus if feature F_1 has an equivalent feature F_2 in Polish-Finnish clauses, F_2 can be considered a functional equivalent of F_1 only if F_2 is an equivalent of F_1 also in Finnish-Polish clauses.

5.2.3 Size of corpus

The next important question concerns the size of the corpus. The four relevant criteria, which I address in turn, are practicality (time constraints, see Reppen 2010: 32, representativeness, availability of parallel texts and their symmetry with regard to external norms.

To start by considering practicality, two issues must be mentioned. First, corpora employed in quantitative studies of grammatical devices can be relatively small in comparison to the lexical studies, where the frequency with which an item occurs is counted, or studies of rare syntactic phenomena (e.g. Huumo et al. 2017 and Jurkiewicz-Rohrbacher et al. 2017). This is because, as noted by Hakuli-

[1]The impact of interference from the source language is separately examined for some linguistic features in relation to the obtained results in Appendix C. Nonetheless, the present study does not aim to be a contribution to the translation studies. From the perspective of contrastive linguistics, the solution of bidirectionality is sufficient.

nen et al. (1980), the point after which more data does not bring much new insight is reached quite quickly in the case of categories occurring frequently.[2]

Second, corpora that need extensive manual annotation are necessarily small, as noted by McEnery et al. (2006) and Reppen (2010), since this work is time consuming. For example, the corpus used in Xiao & McEnery (2006) comprised 100,170 English words and 192,088 Chinese characters, while the corpus used to study temporal notions in Santos (1996) consisted of 50,000 words in English and 50,000 in Portuguese. Tthe final data set for the current study was collected in two stages. First, parallel texts comprising 125,000 Polish words and 105,000 Finnish words formed corpus_1 collected according to the convenience principle, that is, in an electronic format from the Internet, or in rare cases of literary texts turned to the digital form with an OCR software. As the text types show certain asymmetries, for example, in genre, register, or style (explained in the next section), the analysed clauses were chosen to cover broad as possible, but a sufficiently representative, spectrum of different temporal structures (see Section 5.5), forming corpus_2 of (900 Finnish, 900 Polish, total 1800 clauses).

5.3 The asymmetry of texts

The main problem of parallel corpora is the scarcity of bilingual texts and their asymmetry with regard to external norms or factors such as REGISTER, the similarity between texts appearing in the same social context (e.g. law, academic, politics) and GENRE, similarity in formal discourse structure (e.g. letter, leaflet, novel, instruction, news). As noticed by Čermák & Rosen (2012: 413–414):

> "the sum of available translations from one language into another represents the sum of standards of accumulated interest of one culture in another through its texts. The interest may be historically conditional or general and long lasting, covering a well-defined period of time. This fluctuating influence of external factors is particularly significant when comparing the sum of what has been translated between two languages with a smaller number of speakers."

In the following sections, I discuss some external, socio-economic factors that explain the availability of Polish-Finnish parallel texts and their asymmetric distribution across different genres and registers. Subsequently, I argue that external

[2]However, quantitative discourse studies show a correlation between certain types of frequent grammatical feature (such as tense, mood, aspect, person, or number) and the type of text. See Section 5.5.

criteria are less relevant for grammatical studies, in particular those related to temporality.

5.3.1 Size of the speaker population

The first and probably the most important factor that influences the number of produced texts and translations is the size of population of speakers of compared languages. Finland has over 5,500,000 (Tilastokeskus 2014: 2) inhabitants, Poland over 38,500,000 (Nowak 2012: 47). 89.3% of Finland's population are Finnish speakers, 5.3% Swedish speakers and 0.04% speakers of Samic languages (Tilastokeskus 2014: 2).

The biggest groups of foreigners in Finland are Russian speakers (66,379) and Estonian speakers (42,936) (Tilastokeskus 2014: 2). Polish speakers living in Finland are ranked in the 15th place, as the size of the population is between three and four thousand (Tilastokeskus 2014: 8), but there is a relatively big group of temporary workers in Finland, namely 6,500 (Krzywacki & Saarenmaa 2013). The existence of such foreign labour has linguistic consequences, as is shown below.

No precise data is available on Finnish speakers living and working in Poland. The Finnish embassy in Warsaw estimates that there are about 200 Finnish speakers living in Poland.

5.3.2 Economic factors

Poland entered the European Union in 2004. This event had great economic and political significance. The Finnish labour market opened to in 2010. Since then, the volume of Polish temporary labour has been increasing rapidly. Currently it is the second biggest group of temporary workers on the Finnish labour market after Estonians. These workers have been mainly employed in shipyards and construction, to perform jobs which do not require higher education (Hertzen-Oosi von et al. 2009: 36).

As these Polish speakers do not know Finnish or Swedish well enough, the Finnish authorities had good reasons for providing certain materials in Polish translation. Documents issued by the Finnish tax authorities concerning registration of income and payment of taxes can be obtained in Polish. Some tax regulations addressed to Polish companies were also issued in Polish. Additionally, the Finnish Ministry of Employment and the Economy issues and updates

a booklet[3] on working and living in Finland translated into Polish. Due to the small size of the population of Finns living in Poland, the probability of finding corresponding texts translated from Polish to Finnish is rather low.

The second area of sources is trade and investment. Trade exchange with Finland is steady and its value is €2,500,000,000 a year (Pleciński et al. 2014: 38). Polish imports are usually higher than the exports. The main objects of trade are industrial materials and machines. This is reflected in the bilingual websites of Polish and Finnish producers such as Metso or TME, which display marketing materials (product descriptions and catalogues), user manuals, or rules and regulations concerning the purchase and seller's liability.

Although it is possible to find texts translated in both directions in this field, there are distinguishably more texts translated from Finnish than from Polish. This proportion can be explained with the high cost of translation in comparison to the relatively small number of potentially interested investors from Finland.

Polish companies consider the English version of their websites to be sufficient. Nevertheless, this source is quite valuable as the texts vary a lot. One can find extracts marketing materials, press releases, instructions or codes of conduct that vary in style and choice of linguistic structures.

5.3.3 Political factors

Both countries have their embassies which maintain online services in both languages. Public institutions are a good source of bilingual texts, as they may care more than others about the quality of translations.

One could presume that bilateral agreements between states concerning, for example, taxation should be classified here. However, documents of this kind are translated from a third language, mainly English.

Since Poland joined the EU in 2004, many official and legal texts have been translated into Polish. Therefore, EU databases seem to be a promising source of parallel texts. Unfortunately, most legal texts are first issued in English and then translated, so they could not be included in the corpus. The only written texts that are first written in Polish or Finnish are the European Court Judgements (cases where Poland or Finland was the accusing side).

Another potential source of texts related to the EU are transcriptions of discussions at the European Parliament. The number of available source texts thus

[3] https://www.te-palvelut.fi/te/fi/pdf/esitteet/suomeentyohon_puola.pdf

depends on the number of Finnish and Polish Members of European Parliament (13 Finnish members and 51 Polish members) their active participation in the discussions and the choice of language.

5.3.4 Religion

In Poland 86.7% of the population are Roman Catholic (Ciecieląg et al. 2013: 17), while in Finland there are two state churches: Evangelical Lutheran (75.3% of the population) and Orthodox (1.1%). Roman Catholics are a very small minority in Finland (Tilastokeskus 2014: 6). Similarly, only 0.18% of Poles are Protestant. Therefore, it can be barely expected that there would be any mutuality in the Finnish and Polish religious texts.

Despite the size of the community, Poles contribute considerably to Roman Catholic life in Finland. Of the eight Roman Catholic parishes established in Finland two have been run by Poles. Half of working Roman Catholic priests in Finland are Polish. Although some parish websites offer a Polish language version, little content is available in both languages. Surprisingly, a multilingual publication in electronic format handling religious topics is the Polish magazine *Miłujcie się*, of which some issues have been translated into Finnish (obviously in connection with St Mary Parish in Helsinki which is provided as a contact[4]). I do not know of any Finnish religious texts translated into Polish.

5.3.5 Culture and science

In general, there are two types of text which societies find interesting to have translated. On the one hand, we can expect texts that explore the foreign culture, on the other hand, one will find texts discussing global or universal topics. The history of Polish-Finnish translations starts in 1880. From that year until 2006, 89 first editions of Finnish literature were published in Polish translation, and 117 titles from Polish literature were published in Finnish (Szal 2013: 277).

Recently, we may observe a boost in the Polish-Finnish translation branch. This is due to well-developed support institutions such as national literature centres (FILI – Finnish Literature Exchange, the Polish Book Institute) and foundations which support the work of translators. Polish readers show particular interest in Nordic crime stories and children's literature, while Finnish readers are mostly

[4] `http://rakastakaatoisianne.org/pg/fi/yhteystiedot.html`

keen on Polish fantasy and literature about Jewish culture. The genre distribution is therefore far from symmetric.

Although Finnish drama has a certain brand presence in the world, Finnish plays have not been performed on Polish stages, or translated, while Polish plays are present in Finland.

With respect to film, a significant number of Finnish films have Polish subtitles, mostly those which were shown at the annual festival of Finnish films in Warsaw. Regarding Polish subtitles, the situation is complicated. There are Finnish subtitles for films with famous Polish directors, such as Krzysztof Kieślowski, Roman Polański or Andrzej Wajda, but their films are often hard to process in linguistic study as the original scripts are usually bi- or trilingual. Many subtitles can be found in the open parallel corpus collection OPUS (Tiedemann 2012).

Most of the Polish scientific works on the Finnish market are historical books concerning twentieth-century Poland, especially the Second World War (concentration camps, the Holocaust) and communism.

Translations of more global or general importance have been published, mainly works by Jerzy Grotowski (theory of theatre), Ryszard Kapuściński (travel reportage), Pope John Paul II and Leszek Kołakowski (philosophy). The Polish market has been more interested in the commentaries on the unique characteristics of Finnish culture and literature. These appeared in the thematic issues of literature journals such as *Literatura na świecie* or *Czas kultury*. Also monographs on Finnish history, history of literature, mythology, or similar were published, mainly in the 1980s and 1990s.

No parallel texts from the domain of tourism could be found. The most obvious reason is that the assumed scope of geographic and cultural knowledge about the native country or region is always broader in comparison to the foreign tourist. Additionally, texts for foreigners are typically written in English, German, Russian or Spanish, but rather rarely in Polish or Finnish.

5.4 The structure of corpus$_1$ in traditional terms

Monolingual, large corpora (e.g. BNC 2007), the Freiburg-LOB Corpus, and multilingual corpora (e.g. Káňa & Pelouškova 2010) are commonly described following the register and genre structure.

Ensuring balance and text variety in respect to register or genre is certainly useful in lexical and discourse studies, but not necessarily sufficient for studying grammatical phenomena, where the focus is placed rather on the variety of

grammatical structures. Moreover, due to the factors shown above, it is nearly impossible to have the same proportion of the same kinds of texts gained from two source languages.

It comes as no surprise that the registers included are rather unbalanced and asymmetric as regards the source language, as presented in Table 5.1.

Register	Orig. lang. Finnish	Orig. lang. Polish
Reportive-journalism	5,431	2,055
Fiction texts	8,132	8,030
Informative, functional-administrative	6,538	1,998
Informative scientific, popular scientific	8,384	13,207
Legal	8,469	11,565
Dialogue scripts	14,657	19,143
Political	157	4,186
Total	51,768	60,184

Table 5.1: Registers included in the corpus_1 and their size in running words

The most striking disproportion concerns the political register. As I explained in Section 5.3.3, most texts from this group were obtained from transcripts of proceedings in the European Parliament where Poland has four times more members than Finland.

This raises the question as to what is relevant and representative for corpus studies of grammatical notions such as aspect. According to Kilgariff et al. (2006: 129) no gold standard exists. Corpus designers try rather to approach a reasonable representation of all available kinds of texts. According to Biber (1993: 245) "there is no a priori way to identify linguistically defined types," but he agrees that using the results of previous and current studies can help assure that the selection of texts is both linguistically and functionally representative. Consequently, I discuss the problem of representation according to linguistic features in the next section.

5.5 The structure of corpus_2 based on linguistic features

The empirical, quantitative studies of Biber (1989, 1995) proved that texts form clusters according to the frequencies of used grammatical features, but those clusters do not correlate with traditional genre or register divisions, but rather cut

across them (Biber 1989: 39). This implies that representativeness in the field of grammatical studies does not require samples arising from all genres or registers, but the variety of grammatical structures should be ensured.

This raises a question: the present work should include a variety of which grammatical structures?

In his qualitative study of grammatical structures in texts, Werlich (1976: 21–22, 39–41) distinguishes five main text types based on the human cognition: *description*, which is related to perception in space, *narration*, related to perception in time; *exposition*, related to the analysis and comprehension of particular concepts, *argumentation* related to judging facts and concepts, and *instruction*, related to planning.

Werlich presents grammatical correlates for each text type. For example, typical features of narrative texts are a high density of what he calls *dynamic verbs* such as physical actions, adverbials and connectors enabling ordering events in sequences such as *first* or *and then*. Imperatives and substitute expressions of politeness are typical for instructions.

Smith (2003) organises written discourse modes very similarly to Werlich into: *narrative*, *description*, *report*, *information* and *argument*.[5] This division is mostly motivated by variation regarding temporal notions. The first three modes concern situations which are temporally located, while the latter two are atemporal.

The similarity of textual units belonging to one mode is motivated by the notion of progression based on different frames of reference. Narrative texts move along the time of narration, while reports order the situation in relation to the time of speech. Although description refers to temporally located situations, temporal progress does not exist, but instead the text progresses in space. Atemporal modes of information and argument are characterised, according to Smith (2003: 31–34), by the metaphorical motion through the text domain organised according to the unique referent salient in each sentence clause (usually as part of a subject or object). For example, in a paragraph describing the species of tiger, most sentences should somehow relate to the tiger, for example, by naming it or its parts, anaphoric pronouns or possessive constructions.

Since the current study deals with temporality, it should consider a variety of temporal features. The division with respect to the temporal structure of text, thus the frequency of particular temporal features in texts, seems most important.

[5] Smith also admits the existence of *instruction* as discourse mode, but this mode is beyond the scope of her study.

5.6 Differences in the temporal structure between text types

5.6.1 The structure of corpus$_2$

For the current analysis, clauses were collected from the texts characterised above and stratified into three groups, of equal size and equal proportion of Finnish and Polish originals, as presented in Table 5.2.[6]

Text type	Orig. lang. Finnish	Orig. lang. Polish
narrative, literary texts (LIT)	150	150
informative (INFO)	150	150
to-be-spoken (TBS)	150	150
Total	450	450

Table 5.2: Number of clauses chosen from corpus$_1$ to corpus$_2$

For several, mostly practical, reasons, clauses were selected manually and not truely randomly. First and foremost, random selection of clauses from the corpus would require an extensive preprocessing of all texts including proper division of texts to clauses, alignment of Polish and Finnish clauses as well as filtering the clauses which satisfy the pre-conditions of the current study. These could have been done only with automatic tools. Considering the current state of the art, in particular in terms of marking clause borders, these were not possible to fulfil. Instead, whenever the collected set of texts was big enough, the text excerpts were chosen randomly (e.g. narrative, literary texts), whenever possible. The manual selection of clauses had the advantage that obviously errounous translations were immediately excluded.

Clauses in the narrative, literary group arise from fiction, informative clauses were collected from news, company websites, and essays; to-be-spoken texts (henceforth: TBS) are dialogues from literary texts, film subtitles and playscript dialogues.

In order to avoid bias related to the fact that the choice of available texts was small, I ensured that clauses from at least two texts per text type per source language combination were included, and that whenever possible, they have different authors and translators (see Appendix B).

[6]Detailed infromation about each text is given in Appendix B.

The detailed descriptive statistics of all studied features of $corpus_2$ are presented in Chapter 7 following the full annotation scheme from Chapter 6 but, considering the representativeness and temporal properties of text types summarised in Table 5.2 it is worth examining whether the pre-selected types really differ with respect to temporal features.

Following Smith's 2003 idea that some text types are temporal and others atemporal, I provide a tentative analysis of the three groups as to differences in the four temporal features which must characterise each aligned clause[7] accrding to the annotation model (see Chapter 6). These features are: temporal quantification (specific, non-specific, pattern, statement as described in Section 2.5.2), Polish and Finnish tense, and PVA.

The analysis is based on the frequencies of these four features in different text types. They are examined using *mosaic plot* where each box shows the frequencies (numbers of occurrences) of the values of features proportionally to their share in $corpus_2$. Since this is count data, the differences in distributions can be tested with the chi-square test (having ensured that data distribution does not violate the prerequisites for the test). The plots were generated with the `mosaicplot` function (Meyer et al. 2017) which allows for showing the Pearson residuals with colors. The plots can be used to easily obtain information on strength and type of correlation between the features in focus and the text types. Grey is used for residuals between -2 and 2 which do not show much deviation from the expected values, the intensity of blue shows the strength of positive correlation, while the intensity of red indicates the strength of negative correlation. The values between -2 and -4 (light red) and between 2 and 4 (light blue) show some correlation, but only residuals over 4 (deep blue) or below -4 (deep red) show very strong deviations from the expected values.

[7] An original clause and its translation.

5.6.2 Temporal quantification

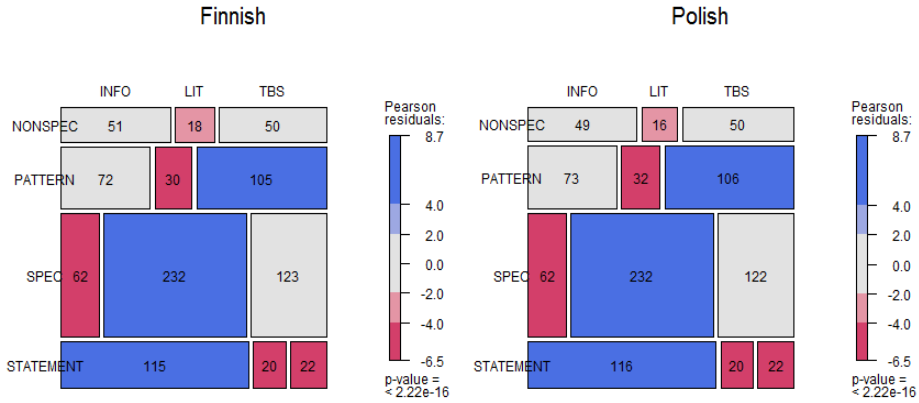

Figure 5.1: Distribution of temporal quantification across text types (left: Finnish clauses; right: Polish clauses)

Figure 5.1 shows that the three samples differ with regard to the temporal quantification of clauses. Although each type of quantification appears in each subset, one type of quantification dominates each type of text. Statements are typical for informative types and atypical for the remaining text types. Temporally specific clauses are most strongly represented in the literary texts, and also the most frequent type in TBS, but not significantly deviating from the expected value. TBS can be characterised by the particularly high frequency of patterns. The frequency of non-specific quantification is quite equally distributed across text types with a light negative residual for literary texts, thus it contributes least to the characteristics of the temporal structure of the chosen samples.

5.6.3 Tense

The difference in the temporal structure of the three text types is even better visible in the comparison of tense use (Figure 5.2). In Polish, the Past tense is predominant in the literary texts, while the Non-past dominates in TBS. Informative texts are balanced in this regard.

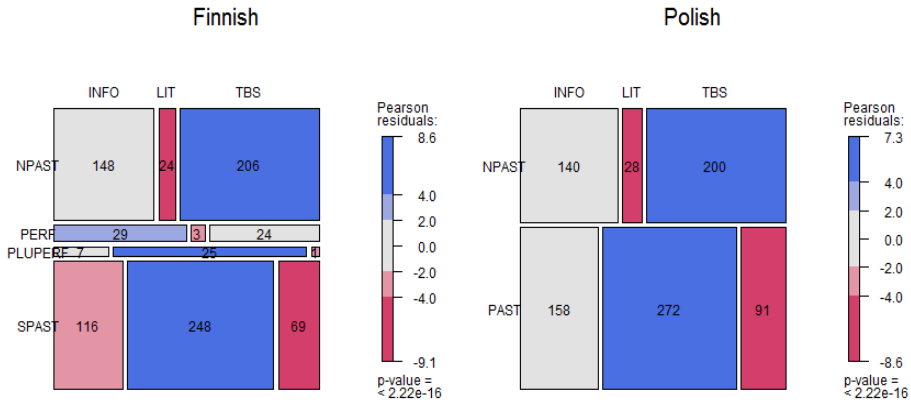

Figure 5.2: Distribution of Finnish tenses (left) and Polish tenses (right) across text types. The Polish Analytical Future was excluded due to infrequency.

Similar tendencies can be observed in Finnish – the Simple Past dominates in literary texts, the Non-past in TBS – but Finnish tense is a four-level variable (see Section 4.4.1). Although the Perfect and the Pluperfect are relatively infrequent, they play a certain role in determining the temporal structure of the three samples. The Perfect is associated with the informative subset and the Pluperfect with the literary texts. Those results sound reasonable, as news often describes situations of current relevance, which is one of the functions of the Perfect, while the Pluperfect mainly fulfils an ordering function. Since the primary narrative line is described in the Simple Past, situations anterior to the main events are described in the Pluperfect.

5.6.4 PVA

The contribution of PVA (see Figure 5.3) is not as clear as in the case of temporal quantification and tense.

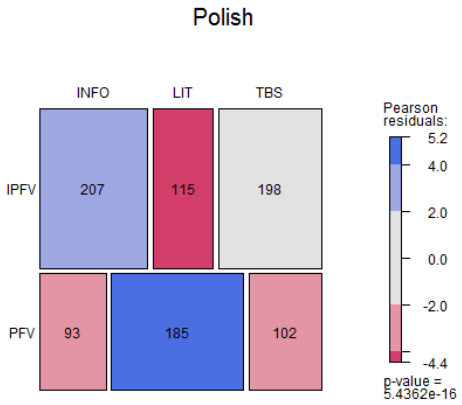

Figure 5.3: Distribution of PVA across text types of the sample

The generally dominated type of PVA is IPFV. PFV is dominant only in the literary texts. In the informative subset and in TBS the proportion is exactly opposite, but the Pearson residual only crosses the value 2, which indicates correlation, in the informative text type.

Thus it seems that PVA's contribution to distinguishing the temporal structure of text types is primarily relevant for the division literary texts versus other text types.

Chapter 6

Annotation scheme

6.1 Introduction

The raw corpus presented in the previous chapter is supplemented with interpretative, linguistic information called ANNOTATION. Linguistic labels covering a wide variety of language phenomena are used in the texts, forming a structure which can be transformed into matrices on which I perform all further computations (see Section 7.3).

In this chapter, I explain the main principle of the annotation scheme – the cross-linguistic comparison (Section 6.2) – and discuss in detail the annotation levels and categories, as well as technical solutions concerning the scheme implementation.

6.2 The issue of comparison in the annotation

Tthe existence of language universals and the cross-linguistic categories has been a matter of heated debate among linguistic scholars for a long time.

While some scholars (Dahl 2012; Dixon 2010; Newmeyer 2007) argued strongly in favour of categories which can be identified across languages, others (Croft 2001; Dryer 1997; Haspelmath 2010, 2016) deny universal grammatical concepts. Instead of pre-established categories, they agree that language-specific categories may exhibit similarities and that functional, cognitive, and semantic explanations for these similarities can be found.

Therefore, one possible method of conducting cross-linguistic, comparative studies is applying *comparative concepts*. Haspelmath (2010) defines them as concepts which do not belong to a language structure, but are defined by linguists

for the purpose of cross-linguistic studies, and are established according to more universal concepts, for example, meanings or functions.

One such account is given by Cysouw (2007) who distinguishes between *categories* which are language-specific and *analytical primitives*. The latter are concepts used to describe the meanings and functions of categories when analysing a particular linguistic data set. The similarity between two categories across languages can be assessed by examining the level of agreement in primitives they both cover. It is important to stress that definitions of categories formed by primitives are not ultimate. Their choice depends on the researcher's hypotheses. It must not be forgotten that the similarity between categories may change, if we approach them by means of a different or wider set of primitives. Due to unforeseeable scientific developments, this assumption can never be rejected.

The descriptions of temporal systems presented in Chapters 3 and 4 show that Finnish does not seem to encode a single linguistic category similar to PVA and covers a similar set of functional-semantic concepts from the field of temporality.

Nonetheless, in Chapter 1 I proposed that some Finnish linguistic categories or combinations of categories do express the same particular functional-semantic concepts as PVA. Therefore, they should occur in parallel texts in certain contexts with a similar frequency as particular values of PVA, or of PVA in combination with some other Polish categories.

In order to discover or reject the possibility of those co-occurrences I examine the set of parallel, annotated clauses in the corpus. The subdomains of temporality explained in Chapter 2, which can be treated as analytical primitives, are realised in clauses by means of linguistic categories related to the field of temporality, as described in Chapter 3 and 4. Additionally, I introduce a set of semantic labels (see Section 6.5.4) for arguments and adjuncts which both serve the purpose of comparison, and shed light on how temporal progress can be mapped to nominal dependents, as suggested in Section 2.6.4.

To do this, clauses must be annotated on two levels – the level of the functional-semantic domain, used for comparison, and the level of language-specific category. The structures from different levels of language should be distinguished too (e.g. morphological units from syntactic units). One possible solution is using a tree structure, as explained in Section 6.3.

In description of the scheme I follow the logic of the tree structure, rather than particular levels of the temporal model as used in the former chapters. Thus, the annotation is presented as follows:

1. the concept of `xml` tree structure (Section 6.3)

2. the annotation of predicates (Section 6.4)

3. the annotation of non-predicative phrases (Section 6.5)

4. the annotation of lexical expressions of temporal localising, temporal durativisation and frequency (Section 6.6)

5. the annotation from the clause and supra-clausal level (Section 6.7).

6.3 Implementation of the annotation scheme

Clauses of corpus$_2$ were encoded in Extensible Markup Language (`xml`), widely used for representing document and data structures. The markup was defined in the self-written DTD-scheme[1] in order to avoid inconsistency in annotation.

The first stage of data processing was to manually extract clauses satisfying the criteria of this study containing exactly one affirmative finite, indicative verbal predicate (see Section 2.1).

The Finnish clauses were parsed with Finnish-dep-par (Haverinen et al. 2014) – an open dependency parsing pipeline, while the Polish sentences were parsed with Wrocław CRF Tagger (Radziszewski 2013). Unnecessary information was then cleaned from CONLL-U format files, which were transformed into `xml` format, aligned and saved as `xml` documents.

Secondly, a semantic layer of annotation was added manually. At the same time, the automatic annotation was post-edited to ensure high quality of the data. I had to perform this work alone due to several constraints. Besides typical time and budget limitations, the process required advanced knowledge of both studied languages. In order to diminish the influence of my own linguistic knowledge limitations, I used several sources. I referred to three Polish dictionaries (Bralczyk 2005; Doroszewski 1969; Żmigrodzki 2007); the Finnish dictionary issued by The Institute for the Languages of Finland (Grönros 2018); two corpora: *The National Polish Corpus* (Przepiórkowski et al. 2012) and *The Language Bank of Finland* available through concondancer Korp (Borin et al. 2012); descriptive grammars of Finnish (Hakulinen et al. 2004) and Polish (Bartnicka et al. 2004; Grzegorczykowa et al. 1998).

As a result, I obtained tree structures with a unique root - `<text>` and its child elements – parallel clauses (tag `<s>`), that is, pairs of clauses where one is the

[1]Document type definition – a reference file where the structure and the legal elements and attributes of a given `xml` document are defined.

original text and the other its translation. Each parallel clause received a unique identification number stored as the attribute `id`. Information about features of each member of a parallel clause (in element `<pl>` for the Polish clause and in element `<fi>` for the Finnish clause) was kept separately. A simplified example of this structure is shown below:[2]

```
    <s id="1579" plstring="Przeczytałam list
samobójczy Virginii wiele razy," fistring="Olen lukenut
Virginian itsemurhakirjeen monta kertaa,">
     <pl>information about the Polish clause</pl>
     <fi>information about the Finnish clause</fi>
</s>
<s id="1580" plstring=" ale wciąż do niego powracam."
fistring=" mutta palaan siihen yhä uudestaan.">
     <pl>information about the Polish clause</pl>
     <fi>information about the Finnish clause</fi>
</s>
<s id="1581" plstring=" Po raz pierwszy zobaczyłam Iana
trzy lata temu w sali wykładowej." fistring="Näin Ianin
ensimmaisen kerran kolme vuotta sitten
yliopiston luentosalissa.">
     <pl>information about the Polish clause</pl>
     <fi>information about the Finnish clause</fi>
</s>
```

Within a monolingual clause, information about temporal quantification, taxis, and clause type were given first (see Section 6.7). Annotation of phrases followed, including the syntactic type of phrase (predicate, subject, object, oblique argument, and adjunct, see Sections 6.4 and 6.5). In the end, adjuncts functional-semantically belonging to the field of temporality received their own annotation according to the subfield they represent (temporal localising, durative temporalisation, type of frequency, see Section 6.6).

[2] A full example is given in Appendix D

6.4 Predicate

6.4.1 Types of predicate

Formally, I distinguish four types of predicate: simple verbs, impersonals (in Polish represented by forms ending with *-no* or *-to* not inflected for person and number, in Finnish according to the description of Hakulinen et al. 2004: §110) and the Finnish progressive construction with the third infinitive Inessive (see Section 4.5.1). Copulae are treated as a subtype of derivation.

6.4.2 Annotation of simple predicates and impersonals

An example of the Polish imperfective predicate *rozgaduje się* 'become.3SG talkative' looks like this:

```
<plverb>
      <pllemma aspect="imperf" derivation="prefixed">
          rozgadywać
      </pllemma>
      <plform>
            <simple>
                  <num>sg</num>
                  <person>ter</person>
                  <tense>praet</tense>
            </simple>
     </plform>
<pfx>roz</pfx>
<plrefl>ref</plrefl>
</plverb>
```

Polish predicates are lemmatised to their infinitive forms stored in the element `<pllemma>` which receives the attribute `aspect`, expressing the value of PVA, and the attribute `derivation` where the basic derivational types are given: copula, simplex, prefixed, translative, and bare. The name of the prefix is given as a separate element. The Polish reflexive particle is annotated together with the verbal unit to which it belongs. The function of the reflexive particle is stored in the element `<plrefl>`.

Finnish verbal lexemes are annotated according to the particular derivational classes described in Section 4.2.1, that is: transitivisers (causative and attitude), detransitivisers (reflexives, reciprocals, decausatives, passive reflexives, translatives), and two temporal structure modifiers (frequentatives, momentaneous). Lexemes not belonging to any of the above classes are classified as neutral.

Morphological information about verbs in text includes number, person and tense, according to the tag set used in the Wrocław CRF Tagger (Radziszewski 2013) for Polish and Fin-dep-par (Haverinen et al. 2014) for Finnish; impersonal forms include only the tense element.

6.4.3 Finnish progressive construction

The scheme contains a separate annotation of progressive construction (see Section 4.5.1) in the form AUX+INF3 in Finnish. The element `<filemma>` contains information about the infinitive lexeme, while the morphological information is stored in the `<inf3>` element.

```
<fiverb>
      <filemma sem_group="causative">
      todistaa
      </filemma>
      <fiform>
            <inf3>
                  <aux>olla</aux>
                  <tense>Pres</tense>
                  <case>Ine</case>
                  <person>1</person>
                  <num>Plur</num>
            </inf3>
      </fiform>
</fiverb>
```

6.4.4 Nominal phrases in predicates

In the case of copulae, the structure of the nominal phrase of the predicate is stored in the <filemma> and <pllemma> elements. It preserves information about the type of phrase in line with the annotation scheme for non-predicate phrases (see Section 6.5.2) and the case (the Nominative is treated as the default case and is not annotated separately). The Finnish predicate phrase *on hyvä* 'be.3SG good' receives the representation:

```
<fiverb>
      <filemma derivation="copula">NP</filemma>
            <fiform>
                  <cop>
                        <num>Sing</num>
                        <person>3</person>
                        <tense>Pres</tense>
                  </cop>
            </fiform>
</fiverb>
```

6.4.5 Free ad-verbals

Separate lexical units with free ad-verbals (see Section 4.2.1) present a big challenge regarding the annotation of Finnish verbs. Therefore, the annotation scheme assumes the separate ad-verbal element <bounder> as in the example:

```
<bounder aggl="ill" syn="free">esiin</bounder>
```

embedding the information about ad-verbal morphology (attribute aggl) and syntactic position (attribute syn).

6.4.6 Light verbs and idioms

The next problem for annotation is caused by strong collocations with verbs, that is, idiomatic expressions and the *light verbs* constructions of the form: verb + noun phrase, such as the Polish *wpuścić w maliny* literally meaning 'let somebody

into raspberries', that is, 'to fool', or the Finnish *saada aikaan*, meaning 'to cause, create'.

As a solution, I annotated them within the verb element and tagged the collocate as a separate element <plfraz> or <fifraz> with a semantic label (see Section 6.5.4) of an argument or adjunct which it replaces, as in the expression *dochodziły do głosu*, literally 'sound.PST.3PL', idiomatically 'become.PST.3PL apparent':

```
<plverb>
  <pllemma aspect="imperf" derivation="prefixed">
            dochodzic
      </pllemma><base>chodzic</base>
      <plform>
            <simple>
                  <num>pl</num>
                  <person>ter</person>
                  <tense>praet</tense>
            </simple>
      </plform>
      <pfx>do</pfx>
      <plfraz role="direction">
            <collocate>do glosu</collocate>
            <pos>PrepNP</pos>
            <case>do+gen</case>
            <num>sg</num>
      </plfraz>
</plverb>
```

6.5 Non-predicative phrase annotation

6.5.1 Syntactic types of phrase

Following Kroeger (2004: 10–11), the main difference between arguments and adjuncts is that the former are closely related to the meaning of the predicate. Usually the omission of an argument causes a loss as to the meaning of the verbal lexeme, while adjuncts can be freely omitted. Several adjuncts may contribute to the same semantic (e.g. spatial or temporal) domain, while only one argument of a

given semantic type is allowed. Finally, arguments may be selectionally restricted by the predicate, for example, as to animacy or physical properties of substance. Although most theories of grammar limit some sets of semantic labels to arguments and others to adjuncts, the present work does not draw such a distinction.

Arguments are subdivided into subject, object and oblique arguments. The subject is an obligatory element of clause in the current scheme, hence null subjects are annotated either as implied, in the case of subject omission (e.g. two in the second independent clause in the sentence), or empty in cases such as impersonals or sentences expressing natural phenomena.

This solution takes the opportunity to assign a semantic label to empty and implied subjects similarly to explicit subjects.

Regarding objects, I annotate the direct objects (fin. *objekti* and pol. *dopełnienie bliższe*). The Finnish direct object was described in Section 4.3. In Polish, the direct object is the argument which, in active clauses, appears in the Accusative, the Genitive or occasionally in the Instrumental. In the transformation of an active clause to passive, the Polish direct object is transformed to the argument in Nominative:

(90) a. *Policjantka kieruje ruchem.*
policewoman control.IPFV traffic.INS
'A policewoman is controlling the traffic.'

b. *Ruch jest kierowany przez policjantkę.*
traffic AUX control.ADJPAS by policewoman.ACC
'The traffic is being controlled by a policewoman.'

6.5.2 Phrase structure markup

Information about non-predicate phrases is stored in one of four elements according to its syntactic function:

- `<plsubj>` , `<fisubj>` – Polish/Finnish subject
- `<plobj>`, `<fiobj>` – Polish/Finnish direct object
- `<plarg>`, `<fiarg>` – Polish/Finnish oblique argument
- `<pladj>`, `<fiadj>` – Polish/Finnish adjunct

Inside the element, morphological information is stored. Child <pos> provides information on the type of phrase (see Section 6.5.3), and whenever it is applicable it has <case> and <num> siblings providing information on case and number.

Polish and Finnish cases understood as suffixes are not comparable. In the annotation scheme, the element <case> is reserved for annotating the grammatical-semantic structure of the nominal phrase. Apart from a suffix, this can be realised by postpositions or a combination of a postposition and a case suffix. For example the bold phrase in (91):

(91) *a społeczeństwo składało się tu od*
and society from.~~PFV~~.lay.IPFV.PST.NEU.3SG REFL here from
zawsze ***z Mattich-typowych-Finów***
always of Matti.PL.GEN.typical.PL.GEN.Finn.PL.GEN
'and the society has always consisted here of Mattis – typical Finns' (S85)

In markup this receives the structure:

```
<plarg role="content" humscale="real" spectype="nonind">
      <pos>NP</pos>
      <case>z+gen</case>
      <num>pl</num>
</plarg>
```

The attribute role contains the semantic label from the set, the attribute humscale and spectype concern information about quantification over referents, as explained in Section 6.5.4.

6.5.3 Basic types of phrase

The basic types of head words in non-predicative phrases are:

1. N – noun
2. Adv – adverb
3. Pron – pronoun
4. Quant – quantifier

5. Num – numeral

Pronouns are divided into: demonstrative (`Dem`), reflexive (`Ref`), indefinite (`Idf`), possessive (`Poss`) and relative (`Rel`) pronouns. The quantifiers marking indefinite quantity are marked separately as `IdfQuant`.

The annotation scheme accounts for the complex character of phrases (mainly of noun phrases) as follows:

1. `GenNP` – noun phrase with non-animate possessor e.g. *flaga Finlandii* 'the Finnish flag'
2. `AdjNP` – noun phrase with some attributive e.g. adjective, prepositional phrase, relative clause
3. `DemNP` – noun phrase with a specified type of pronoun, here demonstrative
4. `QuantNP` – noun phrase with a quantifier, including numerals

Additionally subordinate clauses are marked as *sc*, while implied or empty subjects (in rare cases, other arguments) receive their own marker in the `<pos>` element.

6.5.4 Semantic information about arguments and adjuncts

Semantic labels

In the matter of semantic representation, the model is strongly influenced by functional approaches, mainly the Functional Grammar of Dik (1997) and Deep Case approach of Fillmore (1968). These theories of grammar served as a basic source of semantic labels which are assigned to arguments and adjuncts in the present study. The labels are represented in `xml` as attributes called `role` which belong to the elements containing information about arguments (see Section 6.5.2). The labels themselves should not be equated with semantic roles within any particular linguistic theory.

Since the present study requires bottom-up solutions, it is hard to decide on the correct set of labels in advance.The notion of theme-to-event homomorphism (Sections 2.6.4 and 4.3) is crucial in this decision. Thus, in the corpus I try to identify those arguments and adjuncts which contribute in Polish and Finnish to the notion of measuring out situation.

During the annotation process the set of basic labels was supplemented with necessary additional labels. The description below presents the full set of labels

used in the tagging process. The use of labels is not limited to any particular syntactic type, thus the same label can be used for both arguments and adjuncts. The relevance of particular types is discussed Chapter 7 together with other results.

Qualitative or quantitative change measured on participants

The basic participant distinction concerns AGENT – the causer of some quantitative or qualitative change – and PATIENT. The QUANTITATIVE patient is either coming to existence as in (92) or ceasing to be as in (93). The QUALITATIVE patient exists, but some quality is subject to change as in (94). Finally MEDIA, as in (95), are not subject to change, but they have some property which can be used to measure progress in processing them.

(92) a. *Po dwóch latach [TME*$_{\text{AGENT}}$*]*
after two.PL.LOC year.PL.LOC TME
stworzyła [swój pierwszy
from.PFV.create.~~IPFV~~.PST.F.3SG own.ACC first.ACC
katalog$_{\text{QUANTITATIVE PATIENT}}$*].*
catalogue.ACC

b. *Kahden vuoden jälkeen [TME*$_{\text{AGENT}}$*] loi [ensimmäisen*
two.GEN year.GEN after TME create.SPST first.GEN
luettelonsa$_{\text{QUANTITATIVE PATIENT}}$*]*
catalogue.GEN.POSS.3

'After two years, TME created its first catalogue.' (S1029)

(93) a. *[Alkoholi*$_{\text{AGENT}}$*] liuottaa [sen*$_{\text{QUANTITATIVE PATIENT}}$*] pois.*
alcohol dissolve.3SG it.GEN away

b. *[Alkohol*$_{\text{AGENT}}$*] [go*$_{\text{QUANTITATIVE PATIENT}}$*]*
Alcohol he.ACC
rozpuści.
dispersion.PFV.let.~~IPFV~~

'Alcohol will dissolve it.'(S907)

(94) a. *Przygaszono [światła*$_{\text{QUALITATIVE PATIENT}}$*].*
at.PFV.extinguish.~~IPFV~~.IMPS light.PL.ACC

b. *[Valoja*$_{\text{QUALITATIVE PATIENT}}$*] himmennettiin.*
light.PL.PAR dim.PASS.SPST

'The lights were dimmed.' (S1238)

(95) a. *Olen lukenut [Virginian itsemurhakirjeen*$_{\text{MEDIUM}}$*]*
AUX.1SG read.PSTAP Virginia.GEN suicide.letter.GEN
monta kertaa
many.PAR times.PAR

b. *Przeczytałam [list samobójczy*
through.PFV.read.~~IPFV~~.PST.F.1SG letter.ACC suicide.ACC
Virginii$_{\text{MEDIUM}}$*] wiele razy*
Virginia.GEN many time.PL

'I have read Virginia's suicide letter many times.' (S1579)

In situations where the choice between agent and patient has to be made, for example, in constructions with reflexive markers, the patient label is used.

Change of location

Some situations involve the change of location or position of the MOVER, which may be caused by a POSITIONER along some PATH:

(96) a. *[(Beatrycze)*$_{\text{POSITIONER}}$*] Wiodła [go*$_{\text{MOVER}}$*] [po*
Beatrix lead.IPFV.PST.F.3SG he.ACC on
labiryntach zaświatów$_{\text{PATH}}$*].*
labyrinth.PL.LOC the.beyond.GEN

b. *[Hän*$_{\text{POSITIONER}}$*] johdatti [tätä*$_{\text{MOVER}}$*] [tuonpuoleisen*
s(he) lead.SPST this.PAR the.beyond.GEN
labyrinteissa$_{\text{PATH}}$*].*
labyrinth.PL.INE

'She led him along the labyrinths of the beyond.' (S1262)

The motion's trajectory can be also specified by its starting point – SOURCE, end point – DIRECTION:

(97) a. *[Suomi*$_{\text{MOVER}}$*] muutti [kylistä*$_{\text{SOURCE}}$*]*
Finland move.SPST village.PL.ELA
[lähiöihin$_{\text{DIRECTION}}$*].*
suburb.PL.ILL

b. *[Finlandia*$_{\text{MOVER}}$*]* *przeniosła* *się* *[ze*
Finland across.PFVcarry.~~IPFV~~.PST.F.3SG REFL from
wsi$_{\text{SOURCE}}$*]* *[na osiedla*$_{\text{DIRECTION}}$*]*.
village.PL.GEN on housing.estate.PL.ACC
'Finland moved from villages to suburbs.' (S81)

Transactions such as selling, buying, giving, or change of ownership are also considered as changing the mover's position, but the RECIPIENT label is often used instead of direction. For example, the label of direction was used in the original Finnish sentence:

(98) *[Koko konserni*$_{\text{MOVER}}$*]* *myytiin* *[Pohjois-Koreaan*$_{\text{RECIPIENT}}$*]*.
whole group sale.PASS.SPST North-Korea.INE
'The whole group was sold to North Korea.' (S938)

while in the Polish translation the recipient label must be used due to the usage of the Dative (99a) form instead of prepositional phrase *do*+GEN (99b):

(99) a. *[Bank*$_{\text{MOVER}}$*]* *sprzedano* *[Korei* *Północnej*$_{\text{RECIPIENT}}$*]*.
bank sell.PFV.PSTIMPS Korea.DAT Northern.DAT
b. *[Bank*$_{\text{MOVER}}$*]* *sprzedano* *[do Korei*
bank sell.PFV.PSTIMPS to Korea.GEN
Północnej$_{\text{DIRECTION}}$*]*.
Northern.GEN
'The bank was sold to North Korea.' (S938)

Cognitive processes

Situations concerning cognitive processes describe the relation between an EXPERIENCER and a STIMULUS. As stated in Chapter 4, no unified mechanism has an application for object case marking in Finnish in such situations. Following Van Valin (1999: 374) I distinguish three types of experiencers: cognisers, perceivers, and emoters.

A COGNISER is the psychological locus of cognitive experience, for example, an understander:

(100) a. *[Ona*$_{\text{COGNISER}}$*]* *zrozumie* *[moją*
she from.PFV.understand.~~IPFV~~ my.ACC
poezję$_{\text{STIMULUS}}$*]*.
poetry.ACC

b. *[Hän*$_{\text{COGNISER}}$*]* *ymmärtää* *[runojani*$_{\text{STIMULUS}}$*]*.
(S)he understand.3SG poem.PL.PAR.POSS.1SG

'She will understand my poetry.' (S1344)

A PERCEIVER is the psychological locus of sensory experience, for example, an observer:

(101) a. *[Lew*$_{\text{PERCEIVER}}$*]* *spojrzał* *[na niego*$_{\text{STIMULUS}}$*]*
lion from.PFV.watch.~~IPFV~~.PST on he.ACC
uważnie.
carefully

b. *[Leijona*$_{\text{PERCEIVER}}$*]* *katseli* *[häntä*$_{\text{STIMULUS}}$*]*
lion watch.FREQ.SPST s(he).PAR
tarkkaan.
carefully

'The lion watched him carefully.' (S1628)

An EMOTER is the psychological locus of emotional experience, for example, a liker:

(102) a. *[Mnie*$_{\text{EMOTER}}$*]* *się* *podobają.*
me.DAT REFL please.IPFV.3PL

b. *[Minä*$_{\text{EMOTER}}$*]* *pidän* *[niistä*$_{\text{STIMULUS}}$*]*.
I like.1SG these.ELA

'I like them.' (S1790)

A special type of cognitive process is communication. The producer of an utterance is a SPEAKER, and what the speaker says is labelled as a TOPIC:

(103) a. *(...)[pojat*$_{\text{SPEAKER}}$*]* *keskustelivat* *kovaan* *ääneen*
boy.PL.NOM discuss.SPST.3PL strong.ILL voice.ILL
[kirjallisuudesta$_{\text{TOPIC}}$*]*
literature.ELA

b. *(...)[chłopcy*$_{\text{SPEAKER}}$*](...) dyskutowali głośno [na*
boy.PL discuss.IPFV.PST.VIR.3PL loud on
temat literatury$_{\text{TOPIC}}$*]*
topic literature.GEN
'(. . .) boys discussed loudly about literature' (S1588)

Relations

Some situations concern neither changes nor cognitive processes, but describe some properties of participants or relations between them. ZERO is used as a label for a participant who has some property, in particular being in a state or condition:

(104) a. *[Erikssonin kontti*$_{\text{ZERO}}$*] on tyhjänä.*
Eriksson.GEN container COP empty.ESS
b. *[Kontener Erikssona*$_{\text{ZERO}}$*] jest pusty.*
container Eriksson.GEN COP empty
'Eriksson's container is empty.' (S913)

The physical LOCATION of the POSITIONED referent is shown in (105)

(105) a. *[Kirkko*$_{\text{POSITIONED}}$*] sijaitsee [Mannerheimin*
church be.located.3SG Mannerheim.GEN
Varsovan-kodin lähellä$_{\text{LOCATION}}$*]*
Warsaw.GEN-house.GEN close.ADE
b. *[Kościół*$_{\text{POSITIONED}}$*] mieści się [nieopodal warszawskiego*
church fit.IPFV REFL not.far Warsaw.ADJ.GEN
domu Mannerheima$_{\text{LOCATION}}$*]*
house.GEN Mannerheim.GEN
'The church is located not far from Mannerheim's Warsaw house.' (S262)

CONTENT and CONTAINER specify the abstract relation of containment, ownership, and possession:

(106) a. *[Kansallistunteesta puhuminen*$_{\text{CONTENT}}$*] kuuluu*
national.feeling.ELA speaking belong.3SG
[hiukan samaan luokkaan$_{\text{CONTAINER}}$*] kuin*
slightly.GEN same.ILL class.ILL as

kansanluonteen selvittely.
national.character.GEN investigating

b. *[Mówienie o cesze narodowej*$_{\text{CONTENT}}$*] mieści się*
talking about feature.LOC national.LOC fit.IPFV REFL
[w tych samych kategoriach$_{\text{CONTAINER}}$*], co*
in those.LOC same.PL.LOC category.PL.LOC as
roztrząsanie narodowego charakteru
spreading national.GEN character.GEN

'Speaking about national feeling belongs to quite the same category as investigating the national character.' (S16)

Free labels

Some labels apply to arguments and adjuncts which may occur in different contexts. A BENEFICIARY is an entity on whose behalf the situation takes place:

(107) a. *i poczęstował [Jerzego*$_{\text{BENEFICIARY}}$*].*
and on.PFV.offer.~~IPFV~~.PST Jerzy.ACC

b. *ja tarjosi [Jerzylle*$_{\text{BENEFICIARY}}$*].*
and offer.SPST Jerzy.ALL
'and he offered (a cigarette) to Jerzy'(S1703)

An INSTRUMENT is the tool or means used in the situation:

(108) a. *että tulit [aamujunalla*$_{\text{INSTRUMENT}}$*].*
that come.SPST.2SG morning.train.ADE

b. *że przyjechałaś [porannym*
that at.PFV.drive.~~IPFV~~.PST.F.2SG morning.ADJ.INS
pociągiem$_{\text{INSTRUMENT}}$*].*
train.INS

'that you came with the morning train.' (S1437)

COMPANY is the label assigned to an additional participant in the situation:

(109) a. *Sinähän kuulustelit [Palon kanssa*$_{\text{COMPANY}}$*]*
you.FOC interrogate.SPST.2SG Palo.GEN with

Halttusta,
Halttunen.PAR

b. *W końcu [razem z Palo*$_{\text{COMPANY}}$*]*
in end.LOC together with Palo
przesłuchiwałaś Halttunena,
through.~~PFV~~.listen.IPFV.PST.F.2SG Halttunen.ACC

'All in all, you interrogated Halttunen together with Palo.' (S1562)

MANNER describes in what way the situation happened:

(110) a. *[bezlitośnie*$_{\text{MANNER}}$*] kontynuował Giugiu*
without.mercy.ADV continue.IPFV.PST Giugiu

b. *Giugiu jatkoi [armottomasti*$_{\text{MANNER}}$*].*
Giugiu continue.SPST mercy.less.ADV

'Giugiu continued mercilessly.'(S1303)

REFERENCE is the label characterising a feature or quality assigned to a participant:

(111) a. *Vaakunatuntijat myös pitävät [Suomen*
coat.of.arms.expert.PL.NOM too keep.3PL Finland.GEN
kansalliseläimenä$_{\text{REFERENCE}}$*] leijonaa.*
national.animal.ESS lion.PAR

b. *Heraldycy również uznają lwa [za zwierzę*
heraldic.PL also at.~~PFV~~.know.IPFV.3PL lion.ACC as animal.ACC
symbolizujące Finlandię$_{\text{REFERENCE}}$*]*
symobolising.ACC Finland.ACC

'Heraldics also acknowledge the lion as the Finnish national animal.' (S4)

DEGREE specifies some grade or measure characterised in the situation:

(112) a. *että eräänlaiset metsäläispiirteet korostuvat suomalaisilla*
that some forest.features be.emphasised.3PL Finn.PL.ADE
[vielä enemmän kuin ruotsalaisilla$_{\text{REFERENCE}}$*].*
even more than Swede.PL.ADE

'that some features typical for forest folks are even better visible in the case of Finns than Swedes' (S50)

b. *że niektóre rysy typowe dla ludów leśnych*
that some.PL feature.PL typical.PL for folk.GEN forest.GEN
dochodzą do głosu [wyraźniej$_{\text{DEGREE}}$*] u tych*
to.walk.IPFV to voice.GEN more.sharp at those.PL.GEN
pierwszych.
first.PL.GEN

'that some features typical of forest folks are better visible in the case of the first group.' (S50)

CIRCUMSTANCE is the factor explaining the reason or the purpose of a situation:

(113) a. *Tyttö heilautti [vastaukseksi*$_{\text{CIRCUMSTANCE}}$
girl shake.PST answer.TRANS
nefernefernefermäisiä hiuksiaan.
nefernefernefer.like.PL.PAR hair.PL.PAR.POSS.3

b. *Dziewczyna pomachała [w*
girl po.DELIM.PFV.shake.~~IPFV~~.PST.F.3SG in
odpowiedzi$_{\text{CIRCUMSTANCE}}$ *włosami prostytutki*
answer.LOC hair.INS prostitute.GEN
Nefernefernefer.
Nefernefernefer.

'The girl shook her (prostitute) Nefernefer-like hair as an answer' (S1416)

Quantification over referents

Elements which store information about arguments have two attributes characterising the properties of referents (see Section 2.5.3). Firstly, I distinguish between real (not bigger than a human being or possible to modify instantly) and abstract or functional entities. This is stored as attribute the `humscale`. Secondly, the distinction between individual and generic referents is encoded in the attribute `spectype`. The latter attribute is not obligatory for abstract or functional entities.

6.6 Expressions of temporal localising, durative temporalisation and types of frequency

6.6.1 Temporal localising

Following the descriptions given in Chapters 2, 3, and 4, I distinguish between deictic and absolute temporal localising.

For both types, I use the element <pltemploc>, or <fitemploc>, which may have children: <tu> (deictic localising) or <absolute> (localising to absolute temporal units).

In the element <expression> which is the child of <absolute> three values may appear: UNIT – for expressions referring to canonical temporal expressions such as: date given as the century, decade, year, season, month, day, part of the day or similar; PERIOD – for expressions referring to some non-canonical time units such as childhood, past, or *in one's day*; EVENT – for expressions describing occasions at which the TSIT is localised, for example, Christmas, a conference, or a war.

The attribute relation of the element <absolute> can take the following values: *posterior*, *anterior*, *simultaneity*, *left/right overlap*, *entailment*.

Thus the expression *talvisodan aattona* 'on the eve of the Winter War' is represented as follows:

```
<fitemploc>
      <absolute relation="entailment">
            <expression>unit</expression>
            <case>Ess</case>
            <num>Sing</num>
      </absolute>
 </fitemploc>
```

Expressions localising the TSIT deictically are more complicated to classify. The three semantically homogeneous categories are: units like *yesterday* which cumulates the property of being the canonic temporal expression and relating the TSIT to the TU (where *yesterday* is anterior); periods like *recently* or *in the last years*; DISTANCE where the interval between the end ofthe TR to which the TSIT is assigned and the TU is specified, like *piętnaście minut temu* 'fifteen minutes ago':

```
<pltemploc>
      <tu relation="anterior">
            <expression>distance</expression>
            <case>nom+temu</case>
            <num>pl</num>
      </tu>
</pltemploc>
```

Another important difference between the elements `<tu>` and `<absolute>` are the relations specified by the temporal adverbial. The TR to which TSIT is assigned may be broader than or simultaneous with the TU, but TU cannot be longer than the TR. Thus, the attribute `relation` may not take the value *entailment*. Instead, I introduce the INCLUSION for TR longer than TU. This is particularly useful for disambiguation of adverbs which do not have any other function than marking the relation to TU (labelled together as temporal relativisers). A case in point is the Finnish *nyt* 'now' used in the following sentences from the sample:

(114) a. *Puhun* ***nyt*** *yleisellä tasolla.*
speak.1SG now general.ADE level.ADE
'I am speaking generally.'(S1783)

b. ***Nyt*** *Varsovassa on jäljellä 25 taloa.*
now Warsaw.INE be.3SG trace.ADE 25 house.PAR
'Now there are 25 houses left in Warsaw.' (S858)

In (114a) the TR to which the TSIT is assigned is simultaneous with the TU, while in (114b) it the TR is much broader since the clause describes a current number of houses. Thus *nyt* in (114a) is represented as:

```
<fitemploc>
      <tu relation="parallel">
            <expression>relativiser</expression>
      </tu>
</fitemploc>
```

while the one from (114b) as:

```
<fitemploc>
      <tu relation="inclusion">
            <expression>relativiser</expression>
      </tu>
</fitemploc>
```

6.6.2 Durative temporalisation

Similarly to phrases expressing temporal localising, phrases expressing durative temporalisation are stored as separate elements. Based on the characteristics of durative temporalisation from Section 2.3, their attribute `type` may take the following values: *interval*, *qualitative*, *surrounding*, *left*, *right* or *full* (the latter three refer to boundaries of interval). An example is the Finnish expression *tunnin* 'for an hour':

```
<fidurtemp type="interval">
      <pos>NP</pos>
      <case>Gen</case>
      <num>sg</num>
</fidurtemp>
```

6.6.3 Specifying type of frequency

Expressions specifying the type of frequency in the clause are annotated as separate elements `<fimultass>` or `<plmultass>`, where the attribute `type` takes values: *once*, *summaric*, *specific cycle*, or *unspecific cycle* as described in Section 2.5.4. The children in this element correspond to the phrase description above, as for example the Polish expression *wiele razy* 'many times':

```
<plmultass typemultass="summaric">
      <pos>IdfQuantNP</pos>
      <case>nom</case>
      <num>pl</num>
</plmultass>
```

6.7 Clause properties and supraclausal elements

6.7.1 Sentence clause type

Both dependent and independent clauses were included in the data set. The scheme takes into account that a dependent clause is not only a separately annotated unit, but also that it substitutes some syntactic element of the independent clause. Therefore, the attribute `subtype` of the element `<subordinate>` contains information about the relation between the clauses (attributive, object, taxis, manner, purpose, reason, and condition) while information about the type of conjunction is stored within the element, as shown below, with the example of Polish conjunction *mimo że* 'although':

```
<plclausetype>
     <subordinate subtype="concessive">
     mimo ze
     </subordinate>
</plclausetype>
```

One sentence may also can contain several independent clauses connected with coordinate conjunctions or commas. The first independent clause in the sentence is not annotated. The following clauses receive the element `<coordinate>` in which I mark the type of conjunction.

6.7.2 Temporal quantification

Clauses are qualified with regard to the temporal quantification (specific, non-specific, pattern, and statement, see Section 2.5.2). This information is stored in the element `<plquant>` or `<fiquant>`. The temporal quantification of the sentence:

(115) *Chętnie uczestniczą w nich także goście z*
with.pleasure participate.IPFV.3PL in they.LOC also guest.PL from
krajów nadbałtyckich i Rosji.
country.GEN Baltic.GEN and Russia.GEN
'Guests from the Baltic States and Russia also participate in them with pleasure.' (S847)

receives in xml the representation:

```
<plquant>pattern</plquant>.
```

While the temporal quantification of the sentence:

(116) *W uroczystym otwarciu wystawy*
in festive.LOC opening.LOC exhibition.GEN
uczestniczyli przedstawiciele władz
participate.IPFV.PST.VIR.3PL representative.PL authority.PL.GEN
miejskich Kotki i ambasady.
city.ADJ.GEN Kotka.GEN and embassy.GEN
'The representatives of the Kotka city authorities and of the embassy participated in the festive opening of the exhibition.' (S820)

looks like this:

```
<plquant>specific</plquant>.
```

6.7.3 Taxis

The tactic relation of a clause is stored as the attribute relation in the element <pltaxis> or <fitaxis>. The attribute can take three values: *anterior*, *simultaneity* and *posterior*. The element stores information about the marker of taxis such as conjunction in a dependent clause, or anaphoric expression, similarly to the element expression in temporal localising. For example, the Polish *trzy lata później* 'three years later' is represented as *distance*:

```
<pltaxis relation="posterior">distance</pltaxis>.
```

Chapter 7

Quantitative data analysis

7.1 Introduction

In this chapter, I turn to the data analysis. As described in Section 1.2, I aim to find the correlates of PVA in Finnish and at the same time revise the semantics which lies behind the PFV – IPFV opposition in Polish. I attempt to solve these tasks with quantitative methods.

There are two quantitative approaches to data: hypothesis-testing and hypothesis-generating or exploratory (Gries 2013: 337). The first approach can be applied to data were dependent and independent variables can be easily separated and the hypothesis about the relations is given.

As shown in Chapter 3, the grammatical category of PVA interacts with many temporal categories on different language levels. Additionally, no clear candidates for correlates of PFV or IPFV can be found in Finnish. However, scholars pay attention to DOM, more generally to the argument structure and case marking, and to particular verbal morphemes and lexical temporal and durative adverbials (cf. Nurminen 2017; Tommola 1986). Instead of testing a large set of hypotheses, I now try to answer the following *empirical* research questions:

Q1. What elements of Finnish clauses can be considered correlates of PVA?

Q2. In particular, can Finnish DOM be considered the correlate of PVA in the context of transitive clauses?

Q3. Are the Finnish correlates single features or clusters of features?

Q4. Are the Finnish correlates correlated exclusively with PVA or do they correlate with groups of features including PVA?

Q5. Can the primary semantic correlate of PVA be identified?

I start with the data description presented in Section 7.2. I examine the features

(linguistic categories) included in the annotation scheme designed in Chapter 5 in the light of their frequencies[1] and discuss the extent to which the theoretical description of temporality in Polish and Finnish corresponds with the empirical data.

The distributions of frequencies for members of categories which appear at least 45 times, and thus can be found in at least 5% of clauses, are shown in bar plots. For the Polish data, the distribution of PVA is usually shown in colour. Unless explained in figure captions, the abbreviations are the same as those used in glosses.

Answering empirical questions requires more advanced statistical analysis. Therefore, in Sections 7.3 and 7.4, I examine the data set with distance metrics, hierarchical agglomerative clustering and random forests. I deliver answers to the research questions in the last section. The broader linguistic context is discussed in Chapter 8.

7.2 Features distributions and basic correlation

7.2.1 Predicates distribution in Polish and Finnish

The corpus contains 499 distinct Polish verbal lexemes; copulae were used in 99 clauses. In Finnish, I identified 389 distinct verbal lexemes and 97 clauses with copulae. Four of the latter were aligned with a Polish clause containing PFV. These results are in accordance with conclusions of Biskupska (2018) that the meaning of a Polish verbal lexeme is often more complex than the meaning of a Finnish verbal lexeme, which results in a lack of one-to-one correspondence between the Finnish and Polish verbal lexemes. Instead, the meaning of one Finnish lexeme corresponds with a cluster of Polish, mostly prefixed, verbs derived from the same base.

The full frequencies of verbal lexemes are listed in Appendices E and F. Figure 7.1 shows how many lexemes (abscissa) appeared with a particular frequency (ordinate) in corpus$_2$. Many lexemes are realised only once or twice, while a couple of lexemes have a very high frequency. In order to present them consisely, the frequencies in the plots are normalised with a natural logarithm function, so instead of values between 0 and 100, one can see values between 0 and 5.

[1] In corpus linguistics frequency is a measure of how many times a linguistic feature appears in the corpus.

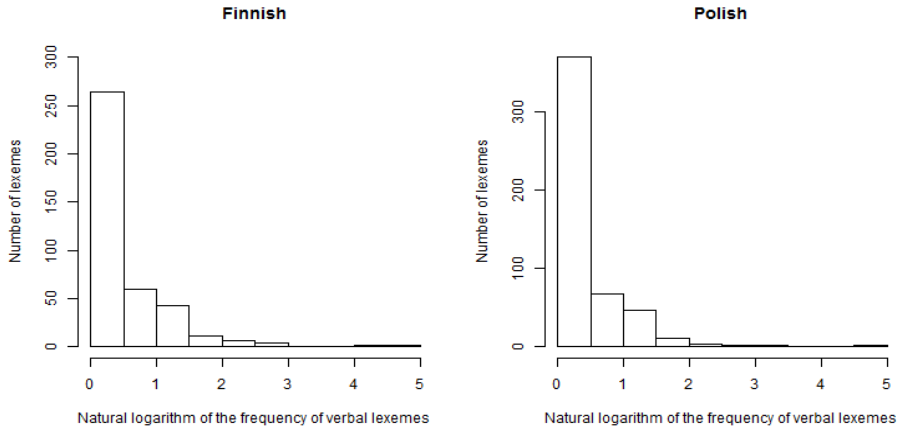

Figure 7.1: Distribution of frequencies of verbal lexemes (normalised to a natural logarithm)

The raw frequencies of Polish and Finnish verbal lexemes have the same quantile structure: the first quartile and median are 1,[2] while the third quartile is 2, but they differ as to the mean. In the Finnish data the mean, 2.87, is greater than the third quartile value, while in Polish the mean, 1.8, is smaller than the third quartile value. That means that lexemes occur only once in the data in Polish more often than in Finnish.

In Polish the five most frequent verbs after a predicative phrase are: *mieć* 'to have.IPFV' (n=32)[3], *być* 'to be.IPFV' (n=15), *wiedzieć* 'to know.IPFV' (n=14), *mówić* 'to speak.IPFV' (n=12), *powiedzieć* 'to say.PFV'(n=12).

In Finnish the five most frequent verbs following the predicative phrase are: *olla* 'to be' (n=84), *sanoa* 'to say' (n=18), *tulla* 'to (be)come' (n=18), *saada* 'to get, receive' (n=14), *tietää* 'to know' (n=14).

The very high frequency of *olla* can be partly explained by the fact that this lexeme is used also for marking the relation of possession or ownership (see Section 4.2.2). Another reason is the higher frequency of light verbs and idioms on

[2]The first quartile cuts the 25% of the data with the lowest values, the median is the number that halves the data, and the third quartile cuts the 75% of data with the lowest values. The mean is the sum of values (here: lexeme frequencies) divided by the number of values (here: the number of lexemes).

[3]The numbers in brackets reflect the raw frequency in the corpus.

the Finnish side – 40 (versus 25 in Polish) where *olla* contributes 26 times, while in Polish *mieć* and *być* appear in such constructions only 8 times. The parallel Polish clauses aligned with the Finnish clauses containing the verb *olla* mostly contain imperfective (77 times out of 84).

Thus, in Finnish, situations are described in 20% of clauses in corpus_2 with the predicative phrase or with the verb *olla*, both strongly corresponding with IPFV.

7.2.2 Morphological properties of verbs

Polish predicates

In corpus_2 IPFV appears in 57% of observations and PFV in 43%.

Figure 7.2 explores the distributions of derivational types and prefixes in the Polish data. As visible in the left plot, the primary derivational types are prefixed and simplex verbs. Imperfectives dominate the simplex types, which is the consequence of the fact that simplex perfectives form a closed group (see Section 3.5.3). Also translatives and bare imperfectives (see Section 3.5.4) appear as marginal notions, while no habitual was observed in the data set.

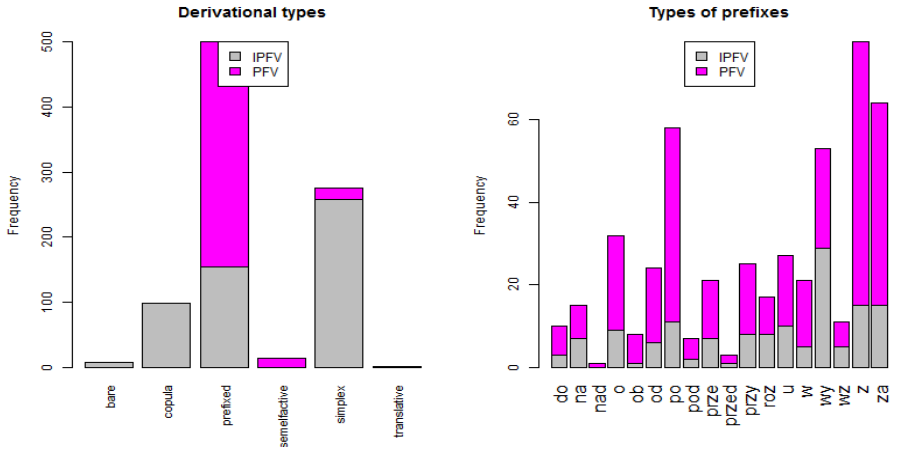

Figure 7.2: Derivational types and prefix distribution in the Polish data

Semelfactives occur only 14 times in the whole corpus_2. Interestingly, they appear only in the literary sample (both Finnish and Polish originals) in the specific

existential quantification. The corresponding Finnish clause contains a momentaneous verb only in three cases.

As visible in the right plot of Figure 7.2, the most frequent prefixes are *z-* *za-* and *po-* which is in line with the prefix distributions observed by other scholars (Grzegorczykowa et al. 1998; Łaziński 2011).

Although prefixes appear mainly in perfective lexemes, the share of imperfectives is 30%. As shown in the right plot of Figure 7.2 not all prefixes show similar "prefectivising strength" – the fourth most common prefix *wy-* 'out' is dominated by IPFV and the prefixes *na-* 'on' *roz-* 'dispersion' and *wz-* 'upwards' are only slightly dominated by PFV.

Further exploration of prefixes shows some moderate associations between three prefixes and text types: *o-* is used more frequently in informative texts, while *z-* is underrepresented in literary texts but overrepresented in to-be-spoken texts. This might be see to supporting claims by some authors that *z-* becomes gradually emptied of its lexical meaning and is prone to the function of a purely aspectual prefix.

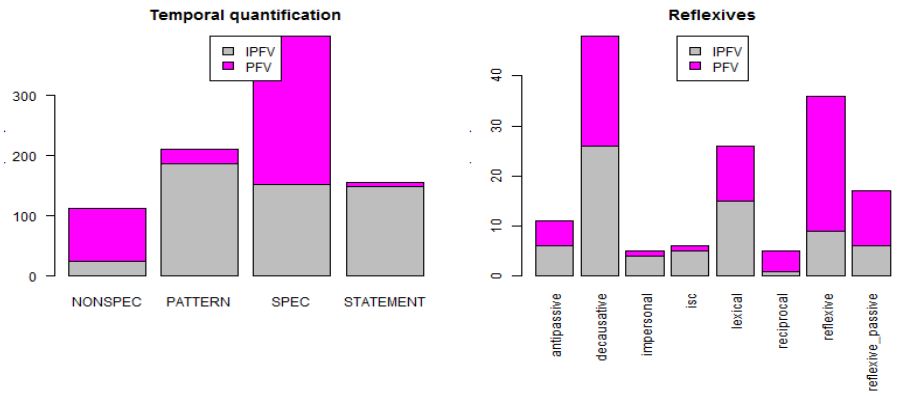

Figure 7.3: The distribution of PVA (left) across types of temporal quantification (SPEC - specific, NONSPEC - non-specific) and across reflexive markers (right).

The left plot in Figure 7.3 shows the distribution of temporal quantification (see Sections 2.5 and 3.8.4) in relation to PVA. In the case of universal quantification, clearly IPFV dominates, while in existential quantification, PFV dominates. Surprisingly, non-specific quantification has a relatively high proportion of PFV.

The right plot in Figure 7.3 shows the distribution of different functions of the Polish reflexive marker *się* (see Section 3.3). The marker occurs 154 times in corpus$_2$ and the scope of its functions in the sample is very heterogeneous. The four main functions of *się* are decausative (n=48), genuine reflexive (n=36), and reflexive passive (n=27). Within decausatives, the distribution of PVA is quite balanced (IPFV: n=26), PFV n=22). Genuine reflexives and reflexive passives are dominated by PFV, which in the case of the genuine reflexives appear three times as often as IPFV, and in reflexive passives twice as often as IPFV. Nonetheless, it cannot be concluded that genuine reflexives tend to be PFV, as this particular type is most unequally distributed across text types. Genuine reflexives arise mostly from the literary sample which has a siginificantly higher frequency of PFV. The lexical reflexive appeared 26 times (11 times with PFV and 15 times with IPFV).

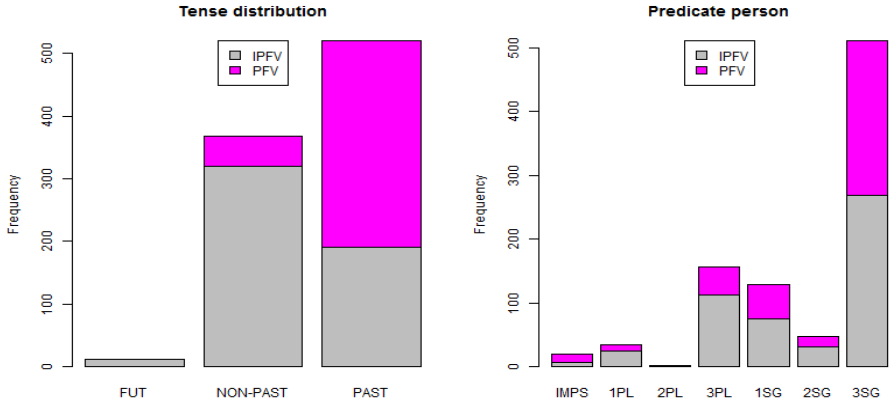

Figure 7.4: Tense, person and number distributions in the Polish data

As visible in the left plot of Figure 7.4, the sample contains few observations of PFV in the Non-past tense (n=50) or the Analytical Future (n=11). Those forms are used predominantly in TBS, and occasionally in the informative sample, but no observation comes from literary texts. This is not surprising, considering that statements with the future temporal reference are generally less frequent than clauses with the past, current or atemporal reference. Also, intuitively, the probability of the future temporal reference is highest in TBS, while the other text types focus often on reporting past events.

A similar, but not unexpected, lack of observations occurs in the case of person

and number of verbs, shown in the right plot of Figure 7.4. Singular is more frequent than plural; In the case of person, the third person is the most frequently used form. Very little data is available for the second person plural. Thus, my data is not balanced in that respect, but rather natural.[4]

Polish idioms and light verbs used as predicates appear 25 times, mostly in informative texts (n=18), in IPFV (n=18), and in the context of universal quantification (n=19). Due to the low number of observations, it is hard to make any statements about relations between that class of predicates and PVA.

Finnish predicates

The distribution of temporal quantification (right plot in Figure 7.5) in Finnish resembles the one in Polish which is to be expected considering that the clauses were translations, so the type of quantification should be kept in most cases.

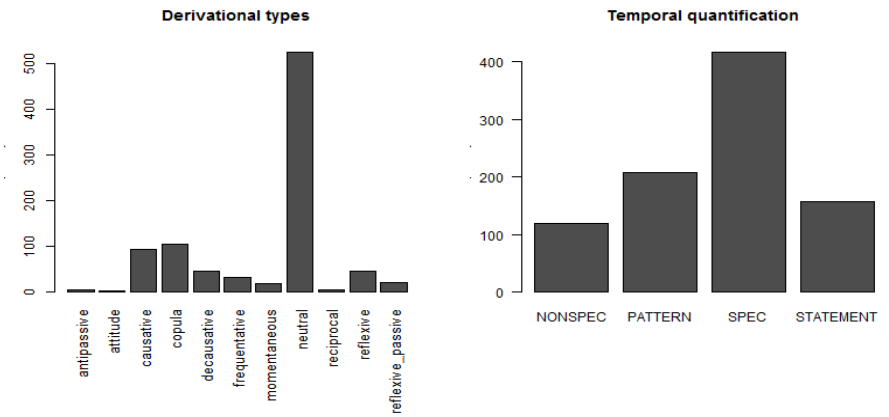

Figure 7.5: Distribution of derivational types and temporal quantification in the Finnish data

In fact, 876 clauses showed the exact correspondence of temporal quantification. The example (117) comes from the fourteen observations where I found no

[4]The best source of data for second person would be spontaneous dialogue, but this kind of data is rarely available in parallel versions. Moreover, the translation would rarely be generated under similarly spontaneous conditions. Authentic, consecutive record, on the other hand, would probably not be natural at all, so parallel versions would not really be comparable.

such correspondence, related to the change of tense in the Polish translation.[5] The usage of the Simple Past in Finnish (117a) suggests that the clause refers to some particular situation temporally localised prior to TU, while the Non-past tense in the Polish (117b) clause suggests the universal quantification.

(117) a. *Mutta lekuri* ***väitti,*** *ettei* *Palo halunnut*
but doctor claim.SPST that.NEG.3SG Palo want.PSTAP
lomalle *vaan pyysi* *rauhoittavia.*
holidays.ALL but ask.SPST calm.down.PP.PL.PAR
'But the doctor claimed that Palo did not want to go on leaves, but he asked for sedatives.' (S1536)

b. *Ale nasz lekarz* ***twierdzi,*** *że* *Palo nie* *chciał* *iść na*
but our doctor claim.IPFV that Palo NEG want.PST go on
urlop, *tylko poprosił* *o* *środki*
holidays.ACC but along.PFV.ask.~~IPFV~~.PST about mean.PL.ACC
uspokajające.
calming.down.PL
'But our doctor claims that Palo did not want to go on holidays,[6] but he asked for sedatives.' (S1536)

The distribution of derivational verbal types in Finnish is presented in the left plot of Figure 7.5. In all, 527 verbal lexemes did not contain any derivational marker. The most frequent are causative affixes which correspond equally often with PFV (n=47) and IPFV (n=46) in the Polish clauses. It seems that derivational transitivisers play a role in aspectual mechanisms of Finnish.

Affixes in the function similar to the Polish reflexive marker appear 123 times, that is, they are less frequent then in Polish. This difference is probably due to the lack of lexical reflexivity in Finnish. Finnish decausatives, reflexives and reflexive passives have a similar frequency as in Polish, but the correspondence to PVA deviates in comparison to Polish in the latter two groups. Finnish genuine reflexives

[5]Such clauses are excluded from statistical models so they do not effect the statistical results (see Section 7.3.3). This is motivated by the fact that if the aligned clauses do not share the type of quantification, the translation is probably far from the original.

[6]This is the Polish translation where obviously the translator got suggested by the Finnish abbreviation of *sairasloma* 'sick leave' to *loma* which usually means 'off from work'. The mistake is visible from the non-idiomatic use of the verb *iść* 'to go' which collocates rather with *zwolnienie* 'to go on sick leave', whereas *wziąć* 'to take' is typically used with *urlop* 'off from work'.

mostly correspond with a Polish clause with PFV, but only each fourth Finnish clause has IPFV counterpart. In the group of reflexive passives the proportion is reversed. Polish clauses with IPFV have 13 Finnish counterparts within this group, whereas clauses with PFV only 8.

Similarly to the Polish semelfactive class, momentaneous verbs appear (with one exception) in literary texts and they always have a PFV counterpart – prefixed or semelfactive.

Due to the lack of frequent one-to-one correspondence it cannot be simply stated that the Finnish momentaneous and Polish semelfactives are full correlates. The semantics of verbs to which they contribute is slightly different. Semelfactives are Perfective as a TSIT can only be assigned to a discrete TR, but the inner situation structure must be non-quantisible 'to produce a quantum of situation'. The last constraint is not the case for momentaneous verbs – it is rather the TSIT, which is so short that it is hardly perceptible as quantisible.

Nonetheless, both groups constitute roughly 5% of the literary sample and represent PFV or the correlate of PFV. Semelfactive and momentaneous morphemes are used in originals and translations.

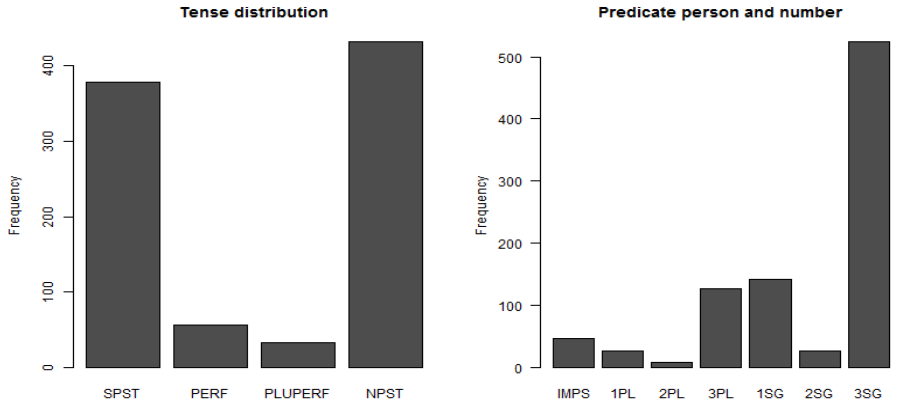

Figure 7.6: Tense, person and number distributions in the Finnish data

As shown in Figure 7.6 all four tenses appeared in the data, though the Perfect and the Pluperfect were rarer than simple tenses. Tense distribution across text types is not balanced (see Section 5.6.3). Additionally, the Pluperfect is used

much less frequently in translation from Polish than in genuine Finnish texts (see Appendix C). The number-person marking is similar to that in the Polish clauses.

The class of free ad-verbals (see Section 4.2.1) appeared only 21 times in the whole data set and it was equally distributed across Finnish originals and translations. This, first of all, means that translators did not try to overuse this class to express the semantics of prefixes or to mark PFV. Ad-verbals appeared in all three text types, mainly in literary texts. Similarly, they predominantly characterised specific situations, but single occurrences of other types of temporal quantification appeared. Bounders appeared mainly in directional cases with the following frequency: Old lative (n=10), Allative (n=2) Illative (n=5), Instructive (n=2), Old separative (n=1), Translative (n=1).

Finnish clauses containing ad-verbals were aligned to Polish clauses containing PFV 19 times, and within this group, only 7 Finnish clauses contained a Total object. IPFV appeared only twice in the context of bounders, both times with the Non-past tense. Once the Finnish clause contained the object in the Partitive, and once in the Nominative. On the basis of these observations it is hard to make any strong statements about the aspectual character of bounders, but they appear as typical of the default past PFV context which can be modified by tense and object case.

Finnish idioms or light verbs appeared in Finnish clauses 41 times as predicates: 13 times they corresponded with PFV, prefixed verbs, 28 times with IPFV (23 times simplex or copula, 5 times prefixed). Within clauses corresponding with Polish IPFV, 20 times predicative constructions contained the lexeme *olla* with noun phrase in the Inessive (n=13), Nominative (n=3), Partitive (n=2), Adessive (n=1), Essive (n=1), respectively. Other verbal lexemes appeared with Illative, Allative, Translative, Genitive and Partitive. While the latter group was used mainly in the context of universal quantification, *olla* was also used in existential quantification.

Interestingly, in the whole sample the 'Finnish progressive' (see Section 4.5) occurred only once. Nevertheless, I cannot conclude that this construction is very rare in Finnish, because it might be typical of the spontaneous spoken language not included in the data.[7]

[7] A quick reference in the Suomi24 corpus of Internet discussions (AllerMediaOy 2014) retrieves e.g. 23,884 forms of *odottamassa* 'wait.INF3.INE'.

7.2.3 Arguments and adjuncts

7.2.4 Semantic labels and cases of arguments and adjuncts

In Finnish, a semantic label was assigned to 900 subjects, 436 objects, 401 oblique arguments and 189 adjuncts, in Polish to 900 subjects, 413 objects, 427 oblique arguments and 191 adjuncts. Finnish arguments and adjuncts were in a form inflected for case in 77%, and in only 67% in Polish.

The most common type of structure in Finnish was two-argument (n=538), followed by one-argument (n=216), three-argument (n=139), and four-argument (n=7). In Polish, two-argument structures are also the most common type (n=488), but somewhat less frequent. One-argument structures are similarly frequent as in Finnish (n=224), while three- and four-argument structures were more frequent (respectively n=157 and n=29).

The distributions of arguments and adjuncts across semantic labels and cases are presented in Figures 7.7 and 7.8. Figure 7.7 shows that same semantic labels are assigned to particular syntactic structures with very similar frequencies in both languages. The three most often used semantic labels are agent, stimulus and mover, followed by location and manner. No label is assigned to all four distinguished syntactic types. In the rare cases of three-way assignments, one of the labels is clearly marginal. It appears impossible that a label could be assigned to both subject and adjunct. The labels zero, speaker, positioned, positioner, agent, and the three types of experiencer are typical of subject. The labels of circumstance, company, location and manner are assigned only to arguments and adjuncts.

The opposite trend in distribution of the labels container and content is related to the separate verbal lexeme of possession in Finnish (see Section 4.2.2).

The combinations of labels are very sparse in both languages. The most frequent type is one argument frame with zero label for the subject, appearing in Finnish 97 times and in Polish 94 times. The second most frequent frame in Finnish is positioned$_{\text{SUBJECT}}$ + location$_{\text{OBLIQUE ARGUMENT}}$ (n=56) and speaker$_{\text{SUBJECT}}$ + topic$_{\text{OBJECT}}$ in Polish (n=47). The third frequent type are in Finnish cogniser$_{\text{SUBJECT}}$ + stimulus$_{\text{OBJECT}}$, and in Polish positioned$_{\text{SUBJECT}}$ + location$_{\text{OBLIQUE ARGUMENT}}$ (n=46). All these frames but speaker + topic are predominantly co-occurring with IPFV.

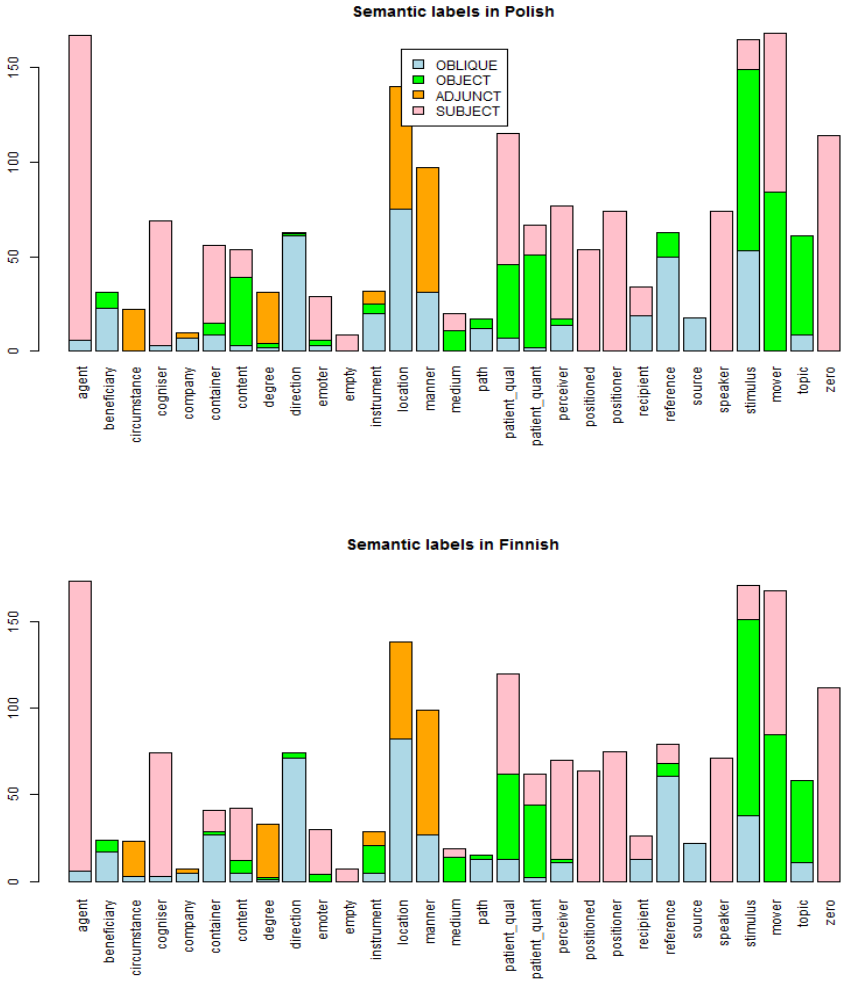

Figure 7.7: The assignment of semantic labels to arguments and adjuncts

Figure 7.8 shows that semantic labels cannot be associated with one case in either language. For example, the label location is marked in Polish mainly with a preposition and the Locative, but also preposition with the Genitive or Instrumental occurs. The colours red and green in the bottom graph are used to mark the typical object cases in Finnish, the Genitive and Partitive. The Partitive has broader distribution than the Genitive, which is in accordance with the theory presented in Chapter 4.

One label used in corpus$_2$ and co-occurring with the Partitive, but not with the Finnish Genitive, does not appear in the literature listed in Section 4.3.2. The instrument, which is usually expressed in Polish with the Instrumental and in Finnish with the Adessive, is expressed in the data with the Partitive. One typical usage of Instrumental-Partitive is in two-argument structures describing movements performed with one body part such as waving one's hand, shrugging one's shoulders as in (118) and (119) below:

(118) a. *Giugiu ugodowo machnął ręk-ą.*
Giugiu conciliatory wave.PFV.SEMEL.PST hand-INS

b. *Giugiu heilautti sovitellen kät-tä-än.*
Giugiu wave.SPST conciliatory hand-PAR-POSS.3

'Giugiu waved his hand in a conciliatory way.' (S1296)

(119) a. *Wolfgang wzruszył rami-onami.*
Wolfgang upwards.PFV.move.~~IPFV~~.PST shoulder-PL.INS

b. *Wolfgang kohautti olk-i-a-an.*
Wolfgang raise.SPST shoulder-PL-PAR-POSS.3

'Wolfgang shrugged his shoulders.' (S1301)

Further case similarities apply to the Allative used in the non-directional function, which has a similar distribution to the Polish Dative. Both cases appear with semantic labels of beneficiary, perceiver, and recipient, which are typical semantic roles of dative use for animate arguments (cf. Næss 2008). I now discuss the semantics of each syntactic type in more detail.

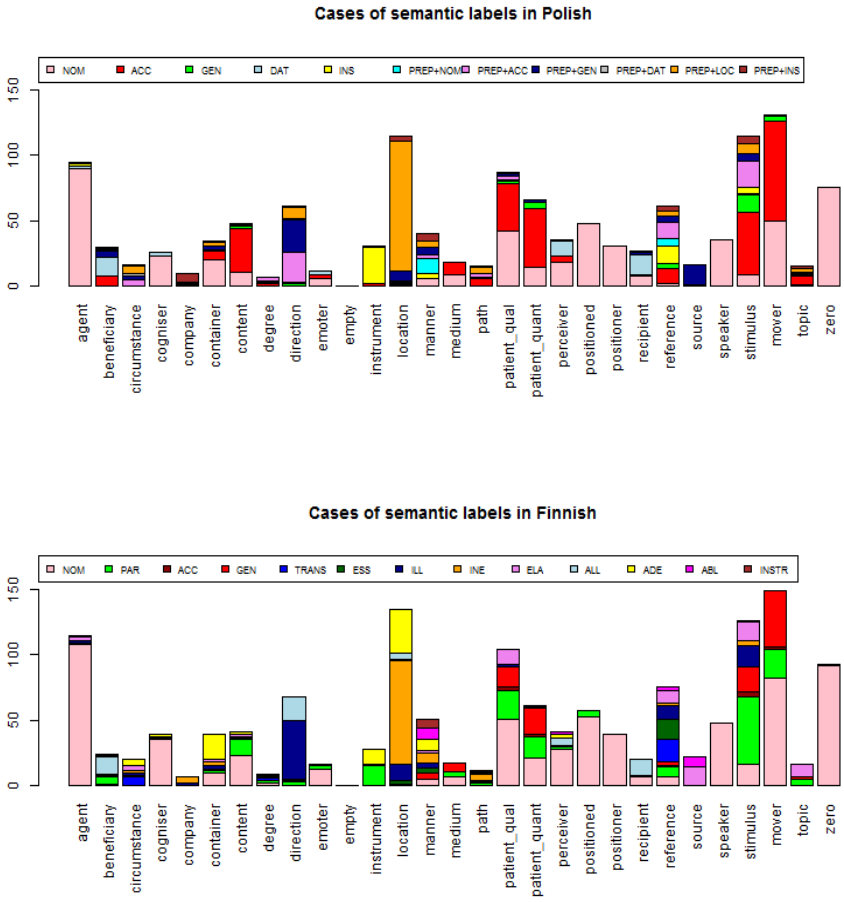

Figure 7.8: The assignment of semantic labels to cases

7.2.5 Subject

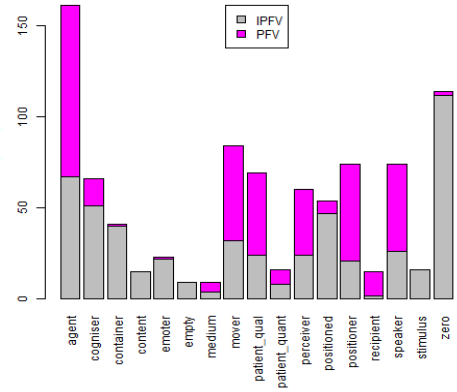

Figure 7.9: Semantic labels of the Polish subject and PVA

Figure 7.9 shows how PVA values distribute across different semantic labels of the Polish subject. The labels zero, stimulus, cogniser, content, container, emoter, empty, and positioned appear mostly in clauses with IPFV. The labels associated with PFV are agent, mover, perceiver, positioner, speaker, recipient and patient. Both types of patient can occur as subject (quantitative patient less frequently than qualitative).

Figure 7.10 shows how the Finnish subject is distributed across different semantic labels. The plot includes the case marking of the subject.

The distribution is very similar to the one for Polish subjects. The Partitive as subject case is used very rarely (n=28), in particular with content (n=8) and reference (n=5). This suggests that the case marking of the Finnish subject cannot play any significant role in theme-to-event homomorphism.

7.2.6 Object

Most of 413 objects in Polish were marked with the Accusative (n=292), the Genitive (n=27) and the Instrumental (n=4). Object was expressed in the form of a subordinate clause 90 times. The Genitive object appeared 8 times in a clause with PFV. In those observations, the Finnish object was marked 5 times with TOT

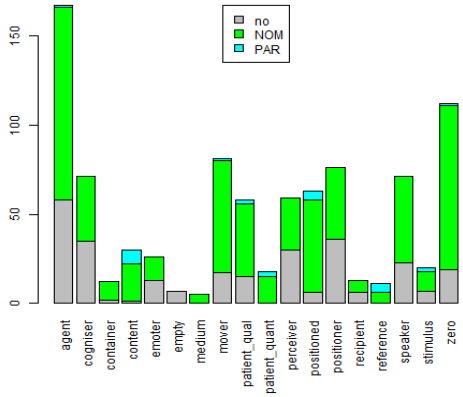

Figure 7.10: Semantic labels of the Finnish subject and case marking

and 3 times with PAR. When the Finnish object was used in clauses aligned to the Polish clauses with IPFV and the Genitive object the Partitive (n=12), no object (n=4) or the Total object (n=3) appeared. The typical semantic labels of the Genitive object were stimulus (n=12), patient (n=5), mover (n=4).

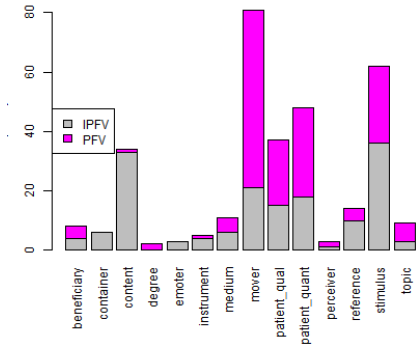

Figure 7.11: The relation between semantic labels of Polish object and PVA

Figure 7.11 shows which semantic labels appear mostly with PFV and which with IPFV. The two labels mostly associated with PFV are mover and both types

of patient, thus arguments associated with the notion of change. The labels associated with IPFV are stimulus and the less frequent labels such as beneficiary, reference, content or container, thus semantically non-prototypical objects, which usually do not undergo any change in the situation.

As mentioned above, Finnish object appears in $corpus_2$ 413 times, that is in 45% of clauses. Nonetheless, only in 320 clauses (35% of $corpus_2$ clauses) it is marked for case: 152 times for PAR 168 for TOT. In the remaining 93 clauses object had a form of subordinate clause.

The object was equally distributed across all three text-type samples in $corpus_2$. Thus, from the discourse perspective DOM behaves differently to PVA which as described in Section 5.6.2 shows difference patterns in different text types. Also the overall proportions are different. While IPFV is the more frequent value of PVA (see Section 7.2.2), TOT dominates in DOM – it appears in nearly 53% of transitive clauses with nominal object.

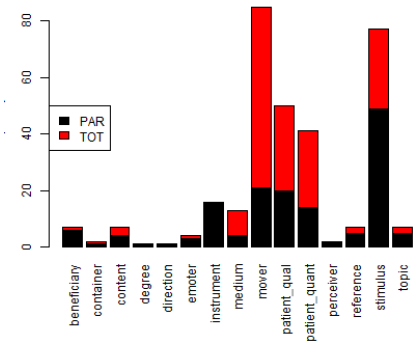

Figure 7.12: The relation between semantic labels of Finnish object and DOM

Figure 7.12 visualises the distribution of DOM across Finnish semantic labels. Here, the distribution is very similar to the one of PVA across semantic labels of object in Polish, in the sense that among the most frequent types, TOT is used most frequently with the labels of mover, patient, while stimulus is mostly marked with PAR. In that respect TOT appears to correspond with PFV, while PAR with IPFV.

The remaining labels are harder to interpret, but their distributions are also less reliable, since they lie below the frequency 20. Nonetheless, both types of PVA

and both types of DOM are to some extent present for nearly all semantic labels.

Oblique arguments

In both Finnish and Polish, only few semantic labels of oblique arguments reach the level of 45.

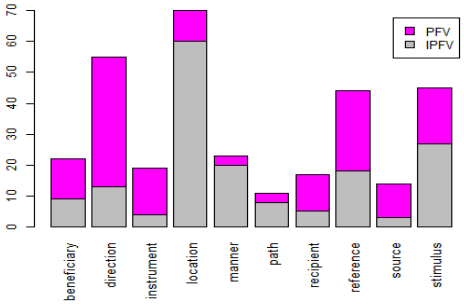

Figure 7.13: The relation between PVA and and semantic labels of oblique arguments

The ten most frequent oblique semantic labels in Polish are presented in Figure 7.13 in the relation to PVA. Location and manner occurs in Polish mostly in clauses with IPFV, direction, source, instrument, beneficiary and recipient with PFV, reference and stimulus have quite balanced co-occurrence patterns with PVA.

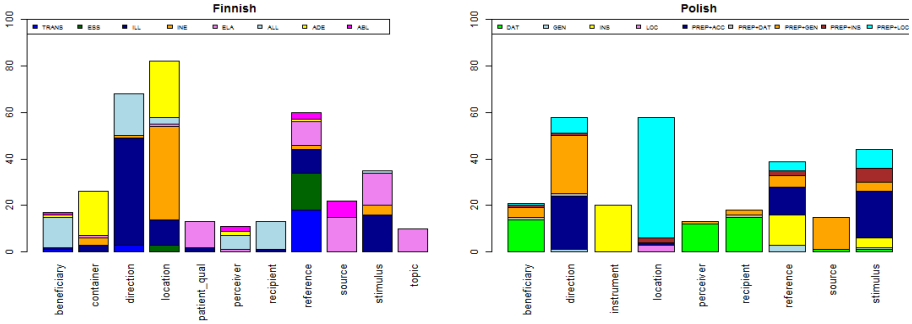

Figure 7.14: The relation between cases and and semantic labels of oblique arguments

Figure 7.14 shows the case-marking patterns for the most frequent oblique arguments in Finnish and in Polish. In Finnish, the most frequent are location (n=82), direction (n=72), and reference (n=61), in Polish location (n=75), direction (n=61), stimulus (n=53), and reference (n=50).

The distributions of arguments marked for case are somewhat different in the studied languages. Direction, location and reference are more frequently used in Finnish than in Polish. In Finnish, each label has one dominant type of marking based on the lative-essive opposition: container, location, and path are the essive types, all other types are lative, but for instance reference has quite some share of the Essive.

7.2.7 Lexical temporal expressions

Temporal localising

Polish temporal localising phrases appeared 89 times in corpus_2 (40 times referring to an absolute unit and 49 deictically) and in Finnish 94 times (50 times referring to an absolute unit and 49 deictically). PFV is more often used with absolute expressions than in relation to deictic expressions (Figure 7.15). The distribution varies more in accordance with tense, as the Polish Past tense occurs mostly with absolute temporal expressions.

In Finnish, the Simple Past also tends to be more frequent with lexical expressions relating to absolute temporal units (Figure 7.16) than to deictic lexical expressions. A similar trend applies to the Pluperfect. Additionally, the Perfect has a considerable share of deictic lexical expressions. Thus, the difference distinguished with grammatical tenses is visible also in their co-occurrence patterns with temporal lexical expressions.

Finally, the differences between the three text types are worth mentioning. Absolute temporal units are most present in the informative texts, and least apparent in the literary texts; in Polish the positive relationship is visible also for TBS. One factor responsible for those correlations is the significant difference in temporal structures of text types (see Section 5.6). Literary texts often have a clear narrative line, where situations follow chronological order. This reduces the need for naming the temporal setting with lexical expressions.

In informative texts, where discourse is organised thematically, but not temporally, the changes in temporal setting are more frequent, which results in a higher density of temporal localising expressions and more mixed use of temporal forms (see Section 5.6.3).

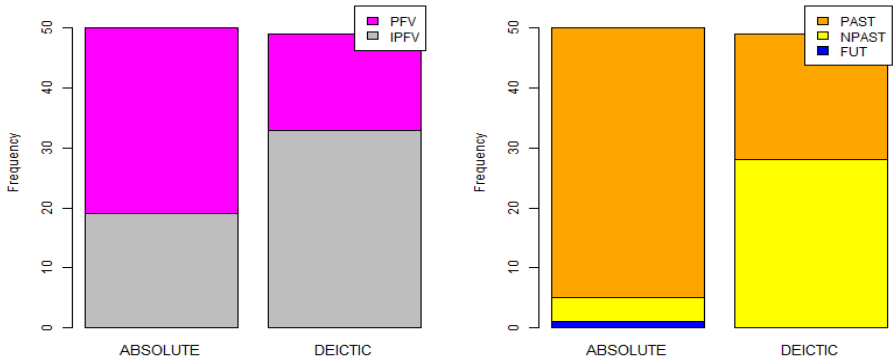

Figure 7.15: Polish temporal localising expressions according to PVA (left) and tense (right)

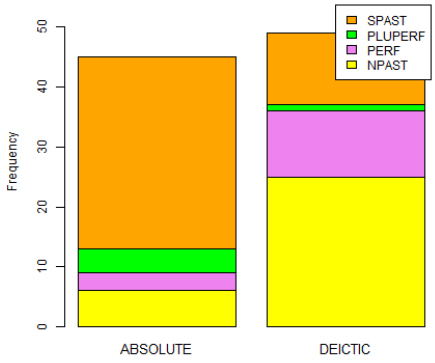

Figure 7.16: Finnish temporal localising expressions according to tense

The increased relevance of specifying the relation between the TR and TSIT is also visible in TBS. Here deictic expressions are more present than in the other two types. On the one hand, the TU is genuine for the time of speaking, and

since most clauses use non-past tenses, there is more need for disambiguation between universal versus existential quantification and past versus future temporal reference than in cases when past tenses are used, in particular in Polish, where the Non-past PFV appears in the function of future tense.

Durative temporalisation

Durative temporalisation (see Section 2.3) was expressed lexically in Polish 25 times and in Finnish 19 times. Durative expressions appeared mostly in the context of IPFV. The few PFV occurrences were related to marking the left or right border of the temporal interval, which characterises how long a situation lasts. The left-bound marker ('since, from') was the most frequent type. Durative temporalisation was expressed lexically mainly in informative texts (Finnish texts n=11, Polish texts n=15), followed by literary (Finnish texts n=5, Polish texts n=7) and very rarely in TBS texts (both in Finnish and in Polish n=3). OSMAs were used only 3 times in the Finnish sample. This suggests that the role of this type of adverbials in expressing IPFV is neglectable.

Type of frequency

Lexical expressions specifying the type of frequency (n=31 in Polish, n=30 in Finnish) are more frequent than those of durative temporalisation, but less frequent than those of temporal localising. Expressions of frequency are distributed in a balanced way across different text types and tenses. The dominating type is the expression of unspecific cycle (n=15 in Finnish original texts and their Polish translations, and n=5 in Polish original texts and 6 in translations). In line with theories outlined in Section 3.8.4), lexical expressions of unspecific cycle appeared in clauses containing IPFV, and in Finnish clauses aligned with Polish clauses containing IPFV. PFV appeared four times, only in the context of summaric expressions or singular occurrence of situation (as also implied in Section 3.8.4). The expressions of unspecific cycles are typical of patterns.

Summing up, in both languages, lexical temporal expressions have similarly low frequency. Since the total number of observations is relatively low, it is hard to make any strong statements. However, based on the differences in their distributions across text types, it can be hypothesised that in both languages, lexical temporal expression plays a similar role in underlying the temporal elements relevant for the particular type of discourse.

7.2.8 Taxis

In corpus$_2$, I identified 130 Polish and 132 Finnish taxical relations of clauses within sentences. Figure 7.17 presents the result for Polish clauses in the context of PVA and tense marking.

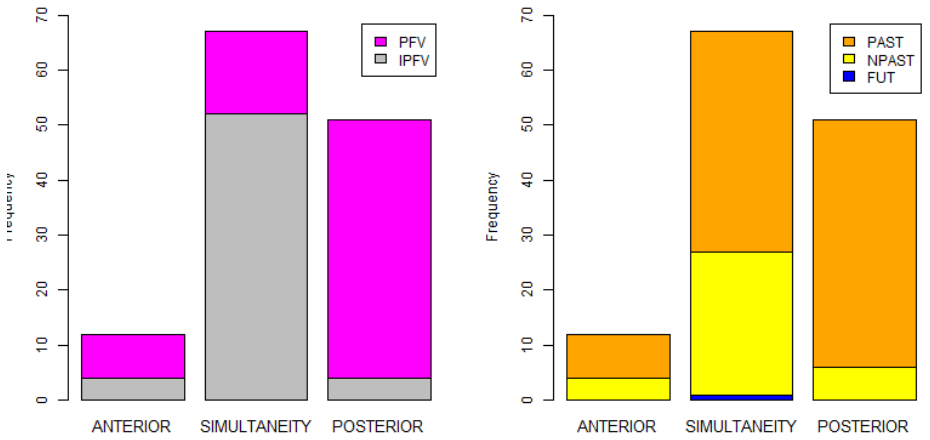

Figure 7.17: Taxis in the Polish clauses according to PVA (left) and tense (right)

In line with previous observations presented in Section 3.8.2, TSIT$_1$ of situation posterior or anterior to some certain TSIT$_2$ appears more frequently with PFV, while TSIT$_1$ simultaneous with some certain TSIT$_2$ appears with IPFV. The Polish Past tense dominates in all three types of relation, but the share of the Non-past in simultaneity is higher than in the non-simultaneous relations.

Taxical relations in Finnish clauses are presented in Figure 7.18. Also here, the Simple Past is used in the context of all three relations, but only in posterior it clearly dominates. Unlike Polish, Finnish has a separate tense for marking anteriority in the past – the Pluperfect. The Perfect is used more frequently in the context of simultaneity of two TSITs than in the context of non-simultaneity. The Non-past tense is also more frequent in the case of simultaneity.

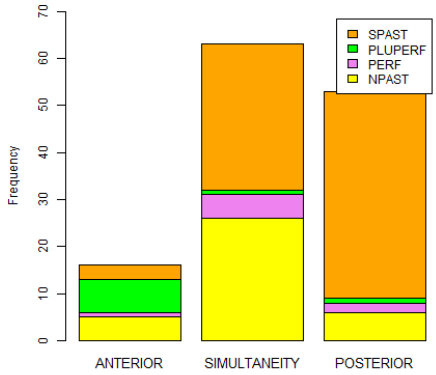

Figure 7.18: Taxis in the Finnish clauses according to tense

7.3 Statistical modelling of Finnish correlates and semantic functions of PVA

7.3.1 Distributional methodology in work with parallel corpora

The main methodological assumption of this study (Section 1.4.2) arises from the *distributional hypothesis* related to the work of Harris (1954), namely, that linguistic elements with a similar distribution in texts belong to the same semantic-functional category (cf. Sahlgren 2008: 33), and the environments of linguistic elements are used as the basis of comparison.

The assumption that distribution correlates with meaning or function has been applied for many years in typology. First, in studies based on multi-lingual questionnaires, as in Dahl's (1985; 2000) research on tense-aspect-mood categories, and recently, thanks to digital progress, with parallel corpora (mainly parallel Bible texts) as in the article by Dahl & Wälchli (2016) on iamitives.

In terms of parallel corpora, the hypothesis must be slightly modified with regard to cross-linguistic comparison. Namely, it is proposed that elements from two language systems but having a similar distribution in the parallel texts refer to the same functional-semantic category.

7.3.2 Distance and similarity

Due to the above, in the present work I attempt to measure the *similarity* between linguistic features based on their distribution. Before moving to measurement, I elaborate more on this approach.

Each clause can be represented as a vector of the following types of elements: grammatical (e.g. the Finnish Past tense marker) or lexical category (e.g. manner adverbial), semantic-functional category (summaric pluractionality) or a combination of both (e.g. subject-patient).

To show how a distribution can be operationalised, let me now take five aligned clauses from the data set:

<s id="1570" plstring=" Wytrzymam ten dzień." fistring="Minä jaksan tämän päivän.">
'I will stand today.'

<s id="1594" plstring=" Ian przysiadł na stole," fistring=" Hän istahti pöydälle,">
'Ian sat down on the table,'

<s id="1595" plstring=" przesunął na bok papiery" fistring=" työnsi paperit sivuun
">
'He moved the papers aside'

<s id="1603" plstring=" Ian poczerwieniał," fistring="Ianin poskille nousi puna,
">
'Ian turned red,'

<s id="1604" plstring=" a jego dłonie rysowały okręgi w powietrzu." fistring="ja hänen kätensä piirsivät kaaria ilmaan. ">
'and his hands were drawing circles in the air.'

For this example I selected eleven features that are encoded for each clause as a Boolean vector:

- F_1 – PFV
- F_2 – IPFV

- F_3 – Polish Past tense
- F_4 – Polish Non-past tense
- F_5 – Finnish Past tense
- F_6 – Finnish Non-past tense
- F_7 – Finnish Partitive object
- F_8 – Finnish Total object
- F_9 – Finnish object stimulus
- F_{10} – Finnish object mover
- F_{11} – Finnish object patient

The realisation of an element in a clause is encoded as one, while the absence of a feature is encoded as zero. In that way, the environments of elements are obtained. Each clause is therefore encoded as in Table 7.1.

Clause id	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}
1570	0	1	0	1	1	0	0	1	1	0	0
1594	1	0	1	0	0	1	0	0	0	0	0
1595	1	0	1	0	0	1	0	1	0	1	0
1603	1	0	1	0	0	1	0	0	0	0	0
1604	0	1	1	0	0	1	1	0	0	0	1

Table 7.1: Encoding features as Boolean vectors

The matrix consisting of row vectors representing all feature environments serves for measuring the similarity between the features through pairwise comparisons. In order to measure the similarity between F_1 and F_{10}, the necessary information can be obtained from the columns F_1 and F_{10} of Table 7.1.

Features	1570	1595	1595	1603	1604
F_1	0	0	1	0	0
F_{10}	0	1	1	1	0

Table 7.2: Pairwise comparison of two features

Let me sum all the cases of environments where F_1 and F_{10} appear together to a number *a*, and the cases where F_1 is present and F_{10} is absent to a number *b*, the cases where F_1 is absent and F_{10} present to a number *c*, the cases where they both are absent to a number *d*.

	F_1 present	F_1 absent
F_{10} present	1	0
F_{10} absent	2	2

Table 7.3: Calculating co-occurrences of two features

This allows me to compute the pairwise SIMILARITY between the features F_1 and F_{10} starting from the general formula (given e.g. in Gries 2013: 341):

$$sim(F_x, F_y) = \frac{a+w_1 \cdot d}{(a+w_1 \cdot d)+(w_2 \cdot (b+c))}$$

where one w_1 and w_2 modify the weights of observations where the factors do not co-occur. Similarity is a measure which takes values between 0 (perfect similarity, two features always occur together and are always absent together) and 1 (absolute lack of similarity, two features never occur together and always one of them is present). In the current study, I use the simple matching coefficient (Rand 1971; Sokal & Michener 1958), where $w_1 = 1$ and $w_2 = 1$, because the absence of features in a sentence is as meaningful as their parallel presence, the equation above reduces to:

$$sim(F_x, F_y) = \frac{a+d}{a+d+b+c}$$

In the case of F_1 and F_{10} this is:

$$sim(F_1, F_{10}) = \frac{1+2}{1+2+0+2} = 0.6$$

The DISTANCE between two features is information about how dissimilar the features are, hence it is the counterpart of similarity:

$$dist(F_x, F_y) = 1 - sim(F_x, F_y)$$

Distance measure based on the simple match coefficient is calculated for each pair of studied features and stored in a distance matrix, analysed in Section 7.3.4.

7.3.3 Data preparation

Some features and observations had to be excluded from the data set used for building the model for reasons such as infrequency or lack of original-translation correspondence. I discuss those problems now.

First of all, I excluded the 14 observations which did not agree as to the type of temporal quantification in Polish and Finnish, as in example (117) in Section 7.2.2.

Secondly, some features were not realised in the sample, mostly because particular syntactic-semantic matches are not possible – for example the semantic label *speaker* was assigned in the Polish corpus only to the subject. All features included previously in the annotation, but never appearing in the data, were not counted.

Finally, the annotation scheme presented in Chapter 6 allowed for a very large number of binary features, but many of them appeared irrelevant from the broader perspective as they are very infrequent (they occurred in less than 40 observations which comprises less than 5% of all observations included in the model). For example, the Finnish progressive occurred only once, so including that in the model should not change the results. The lexical expressions of the type of frequency or durative temporalisation were also too infrequent to be used.

Many features occured infrequently because annotation of their subtypes was overspecific compared to the overall frequency of that category in corpus_2. This applied in particular to the classifications of arguments, taxis, and lexical expressions of temporal localising. In order to preserve some information which otherwise would need to be excluded from the study, I aggregated some features. Thus, the semantic labels of subject and object were aggregated into macrolabels (see Appendix G). The criterion for this procedure was whether object and subject undergo or cause some change in time in the situation.

The nominal elements of light verbs, idioms, and bounders were treated as oblique arguments, otherwise they would need to be excluded entirely.

Oblique arguments were a particularly heterogeneous class, and only a couple of semantic labels crossed the benchmark frequency. Additionally, most semantic types of oblique arguments were marked with several cases, so including features consisting of case and semantic label combinations would not be possible.

However, the semantics of arguments is partly determined by the semantics of subject and object. For example, the semantic label of the subject or object *mover* increases the probability that the oblique argument (if present in the clause) will have a label *source* or *direction*, but at the same time the probability decreases that

the oblique argument will have the label *location*. Additionally, the labels source and direction should differ as to case-marking. Therefore, only the case marking was included for oblique arguments.

Since lexical expressions were quite infrequent, only the main types of taxis relation (simultaneous versus sequence) and the main types of temporal localising expression could be included in the model, but not more fine-grained categorisations. The full list of features included in distance matrix is provided in Appendix G.

7.3.4 Hierarchical clustering methods

Through repeating the computation described in Section 7.3.2 for all pairs of features, I obtain a matrix of coefficients called the DISTANCE MATRIX (in this case a matrix of distances between linguistic features). Due to its size (63x63), the matrix is hard to analyse and interpret by looking at it.[8]

Finding most similar groups of features, called clusters, is not an easy task. I solve it by applying HIERARCHICAL AGGLOMERATIVE CLUSTERING.[9] In this method, all features are treated as individual clusters and those with the least distance are merged, resulting in a binary tree structure.

Each fusion of two features requires updating the distance matrix as the matrix size reduces from nxn to $(n-1)x(n-1)$ such that distances computed for the larger matrix become incorrect. Therefore, we need to recalculate the distance between the remaining clusters and the new fused cluster.

For updating, several methods are standard. The first of these is the *single linkage* method, which uses the minimum distance of clustered elements to remaining features. This approach, however, generally understates the distance of the cluster to the remaining features, as only one element of the cluster – the one with the smallest distance – is taken into account. The second method, *complete linkage*, uses not the minimum but the maximum distance of clustered elements to remaining features. By the same token, this leads to generally overstating the cluster's

[8]However, it is represented as a *heatmap* in Appendix H, where the colour scale is used to represent the value between 0 and 1 in order to easily find the closest and the farthest categories.

[9]The data stored in xml trees were transformed into a binary matrix with self-written Python script. The rest of the computations and visualisations was performed with R software. I created the distance matrix with the `dist.binary` function from the library `ade4` (Dray & Dufour 2007), the only function which allows for flexible choice of binary distance metric. This choice complicates the problem of cluster validation, as the commonly used `pvclust` (Suzuki & Shimodaira 2006) library cannot be used because it does not support binary data with the simple match coefficient.

distance to remaining features, as only the element with the largest distance is taken into account for updating.

The *average linkage* method tries to overcome the problems of the preceding two methods by using the mean distance between the cluster's elements to remaining features. Yet, this method also ignores cluster-internal variance[10] of distance to remaining features when rearranging the distance matrix. This is problematic, since, with rising variance, the precision of all three methods decreases. In extreme cases, this may result in incorrect clustering, especially when the number of features is large.[11]

Therefore, I need to correct for the cluster-internal variance of distance to the remaining element as suggested by Ward (1963). This is done through minimising the sum of squared differences of distances from remaining elements to clustered elements. Hence, Ward's method is more precise than the preceding ones, so I choose it for updating.

Building clusters is not enough, as the structure can be found in any, also random, data. Therefore, the dendrogram must be divided into interpretable, most likely valid, clusters well-separated from other clusters. This can be done with silhouette analysis (Rousseeuw 1987) [12] where, for each object within the cluster, two measures are compared: similarity within the clusters and similarity to the next neighbouring cluster.

In other words, one tries to find out how easily the objects could change their positions from one cluster to another. As a result, the silhouette width is obtained for each cluster solution – a value between -1 and 1. The higher the value, the more likely it is that the cluster is well matched. The authors of the method suggest the value 0.2 as a lower bound. Values below this indicate a lack of structure in clusters.

The three-cluster solution marked with red rectangles in Figure 7.19 is relatively good, as its average silhouette is 0.2483. I now turn to the interpretation of clusters.[13]

[10] Variance is used in statistics to measure how much each observation in the data deviates from the mean.

[11] As a result, early errors would persist through all updates of the distance matrix.

[12] Available e.g. in the R package `cluster` (Maechler et al. 2018).

[13] For the dendrogram used in this study, the two-cluster solution has the highest average silhouette width (0.3319), but it is less insightful than the solution with three clusters.

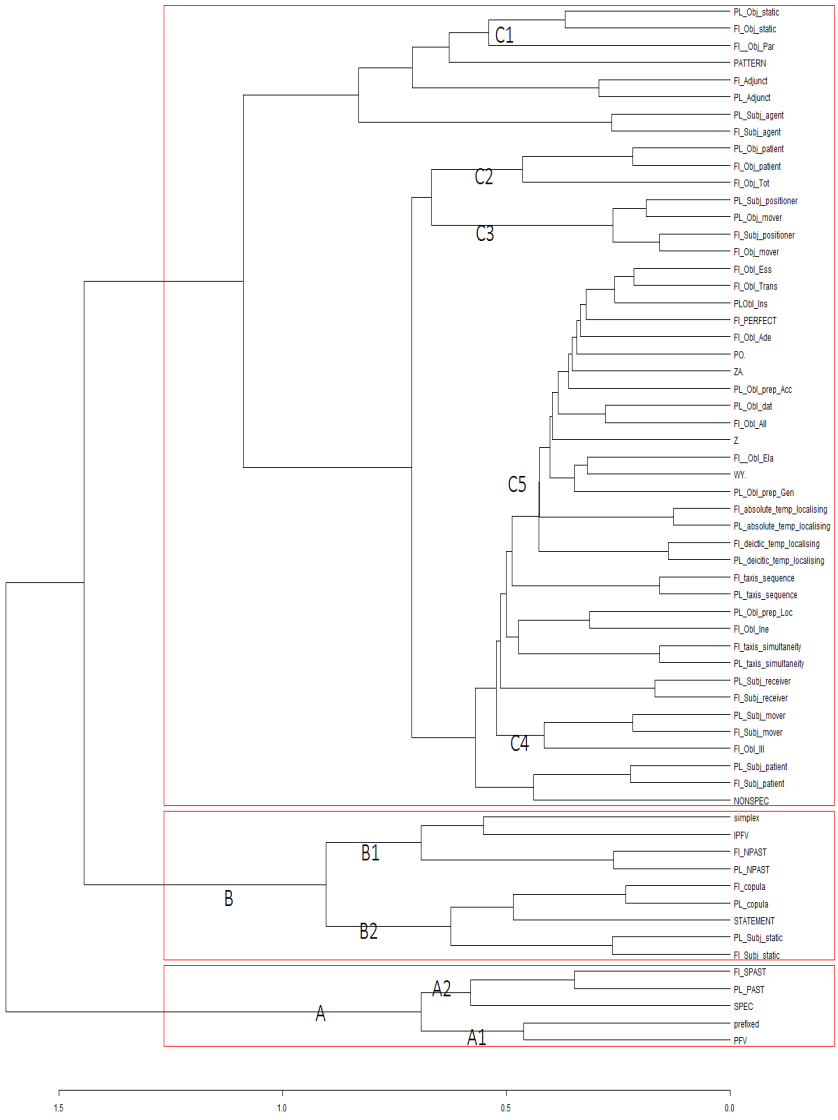

Figure 7.19: Dendrogram visualising the similarity between different features considered to contribute aspectually

The scale next to the dendrogram shows at what level features cluster. The smaller the number, the more similar features are.

Cluster A on the bottom contains PFV and its closest correlates. In Polish PFV is expressed with prefixes, so the relation between these two features is clear and visible in Cluster A1. Cluster A2, consisting of the Polish Past tense and the Finnish Simple Past, correlates with the specific temporal quantification. This cluster is the second correlate of PFV.

Cluster B shows features proximate to IPFV, which clusters first in B1 with simplex verbal forms. This cluster is joined by the non-past tenses. The second branch, B2, consists of copulae clustered with statements, and then with static subjects (labels zero, container, positioned, content, stimulus, emoter, and reference, described in Chapter 6). This result makes sense since the observed copulae are IPFV in Polish.

Cluster C can also be divided, although the clusters are less stable. Cluster C contains the most frequently used prefixes, most types of Finnish and Polish argument, as well as other expressions of space and time parameters in clauses. Although these results are not necessarily reliable, I would treat them as cues for future investigations.

First of all, DOM is more strongly correlated with the class of arguments than with the temporal categories in comparison to PVA. This is not surprising, since contrary to PVA or tense, direct object is not obligatory in each clause. Total object (in Cluster C2) is more closely related to patient objects and positioner subject + object$_{\text{MOVER}}$, while Partitive object (in Cluster C1) is related to static objects (the semantic labels container, reference, beneficiary, perceiver, emoter, topic, stimulus, and content). The Illative does not appear in the cluster with mover objects (C3), but with mover subjects (C4). Patient subjects do not have any close oblique correlate in Polish or in Finnish.

Finally, the differences between the four most frequent prefixes are visible in Cluster C5. The prefixes *po-* and *za-* are close to each other, but it is hard to associate them with any particular element, and therefore explain their meaning or function. The prefix *z-* is even harder to interpret in that respect, but its relation to the other prefixes is very weak. In contrast, *wy-* clusters first with the Elative argument, and then with the prepositional Genitive argument. In other words, the occurrence of *wy-* in the data is more related to the corresponding oblique argument than in the case of other three prefixes. The prefix *z-*, which is related to the preposition *z* 'from' does not join this cluster. This is very much in line with the work of Łaziński (2011) who shows that *z-* behaves quite randomly in

that respect.

Thus, cluster analysis supports the theories that some prefixes are emptied of their spatial, 'prepositional' meaning. On the other hand, the position of *wy-* in the dendrogram shows that although prefixes are closely related to PFV as its marker, they also keep close relation to arguments.

In order to ensure that all important relations of PVA were shown in the dendrogram, I examine the actual values from the distance matrix given in Appendix I.

In general both PFV and IPFV are quite distant from other features. The most similar ones appear first at the distance 0.46-0.56 (values bolded in the table from Appendix I). Of all the closest features, only one is not visible on the dendrogram, because this relation is not symmetric. It is the Total object, which has the coefficient 0.5691. The features closest to the Total object (the semantic macroroles of the object) are in the distance matrix on the level 0.37 - 0.41, so the PFV is already relatively distant. The counterparts do not relate in the same way; the IPFV and Partitive object are quite far from each other. It is important to observe that the whole relation between object and PVA is asymmetric in terms of co-occurrence patterns. Object as category is generally less related to IPFV than to PFV, since the latter has lower coefficients with all types of semantic macrorole of object. Strictly speaking, the same pattern can be observed for oblique arguments – the distance coefficients are generally lower for PFV than for IPFV.

As stated above, Cluster A suggests that (both in Finnish and in Polish) tense is a close correlate of PVA. It could be expected that the relation between IPFV and the Polish Non-past tense should be particularly strong – PFV can never be used in the context where a TSIT is assigned to a TU. Such cases, however, do not dominate in the data: IPFV is used in the context of the situation overlapping with the TU 71 times, while PFV with the Polish Non-past tense in the function of future temporal reference occurs 36 times.

More relevant is temporal quantification – all the compared cases (n=107) relateto existential temporal quantification. However, 261 predicates formulated in the Polish Non-past tense refer to universal temporal quantification, and here IPFV (n=248) indeed dominates over PFV (n=13). This means that PFV is used only in 4% of contexts of universal temporal quantification.

Secondly, the relationship between tenses and values of PVA is symmetric. IPFV is as close to the Non-past tense, as PFV is close to the Past tense, although this result is hard to predict. A glance at Figure 7.4 should suggest that the distance from IPFV to the Non-past would be smaller than the distance from PFV to the Past tense, because IPFV in the Non-past tense looks more dominant than the Past

tense with PFV. Here again, temporal quantification plays a certain role. In the existential temporal quantification, PFV (n=315) is used more frequently than IPFV (n=105). In the context of universally quantified clauses, IPFV remains dominant (n=85) over PFV (n=18). Thus, PFV is used only in 25% of contexts referring to universal quantification. But this means that PFV in a universally quantified context is relatively more frequent in the Past tense than in the Non-past tense.

The overall proportion between universal (n=113) and existential quantification (n=420) is reversed in the Past tense in comparison to the Polish Non-past tense. In the Past tense universal quantification appears only in 21% of the studied clauses (hence: existential quantification is present in 79%), and in the Non-past tense it is in 70% of clauses (hence:existential quantification is present in 30%).

In other words, a three-way interaction is observable. Situations described in the Past tense are more likely to be existentially quantified whereas those described in the Non-past tense are more likely to be universally quantified. At the same time, clauses which contain existentially quantified clauses in the Past tense are more likely to be described with PFV, while universally quantified clauses in both the Non-past and the Past tense are more likely to be described with IPFV. The existentially quantified clauses in the Non-past tense are more likely to be described with IPFV.

7.4 Predicate-argument structure in Finnish and PVA

7.4.1 Random forests

Cluster analysis suggests that the primary function of PVA in the studied clauses is temporal. PFV together with the Past tense is the typical pattern for expressing existential, specific quantification, while IPFV is closely related to the Non-past tense and statements, often expressed in the particular form of copula.

Cluster analysis suggests that the closest correlates of PVA in Finnish are tenses. This correlation is caused partly by the fact that, in the studied indicative clauses, tense and aspect are both related to temporal quantification. In other words, Finnish tenses are correlates of PVA, because PVA is also correlated with tense in Polish. The dendrogram shows that predicate-argument structure has less in common with PVA than tense and temporal quantification, and therefore, the exact relation between predicate-argument structure and PVA is hard to interpret. In particular, the relation between DOM and PVA remains unclear. According to the distance matrix, the Total object is a close correlate of PFV, but this relation is not

well visible in the dendrogram because Total object has even closer, non-verbal correlates.

Some elements could not be included in the model due to their infrequency. Nevertheless, Finnish data can be further explored against PVA with a method called *random forest* invented by Breiman (2001), which was applied to linguistic problems relatively recently by Tagliamonte & Baayen (2012). The computation is possible with the R package `party` (Hothorn et al. 2006a; Strobl et al. 2008).

The method is based on the concept of classification and regression trees (Breiman et al. 1984) where the data is recursively divided with binary splits in order to achieve the most homogeneous groups with respect to the predicted (dependent) variable. A *conditional inference tree* (Hothorn et al. 2006b) is created from a random subset of observations and predictors, and its accuracy is compared against the observations not used for building the tree. Predictors included in the tree are tested for independence, and the most strongly associated variable is chosen on the first node. This process is repeated as long as no statistically significant associated variable can be identified. Random forests are grown from many conditional inference trees. Thus, the prediction of the value of a dependent variable for a given observation happens in a sort of 'democratic election' in the forest.

In random forest, the utility of each predictor can be evaluated with a permutation importance measure, to which I refer as VARIABLE IMPORTANCE. As explained by Tagliamonte & Baayen (2012), the values of a predictor are permuted randomly and assigned to the independent variable. If a predictor is strongly associated with the response, such shuffling should greatly reduce the accuracy of the model. Thus, low variable importance indicates little impact of the predictor on the performance of the model.

As argued by the authors of the method, random forests are an efficient solution for studies where the number of predictors is high in comparison to the number of observations. Through permutation it is possible to avoid bias towards the order of predictors and overfit of the model.

7.4.2 Model 1 – predicting PVA from grammatical features

I assume that if Finnish has some equivalents of PVA it should be possible to determine the aspect values using the features of the Finnish clause. Model 1

consists of purely grammatical Finnish categories,[14] as shown in Table 7.4.[15]

Predictor	Levels	Explanation
Fitense	FI_NPST, FI_SPST, PERF, PLUPERF	Finnish tenses
Fiobject_case	TOT, PAR, NONE	Types of Finnish direct object
Fioblique_case	ALL, ILL, ELA, ABL, INE, ADE, TRANS, ESS, INSTR, LAT, GEN, PAR, no	Cases of Finnish oblique arguments

Table 7.4: Predictors and their levels in Model 1

An example tree is given on the next page in Figure 7.20.

[14]The infrequent lexical elements were omitted, because their contribution to the model is restricted by infrequency of particular types, while the increase of predictors increases the cost in higher computation time. In other words, a single observation of a particular category cannot be more informative for the model than 1/876, thus it will necessarily have a low variable importance.

[15]This is the simplest model: adding any other grammatical elements such as part of speech, number of arguments, or case of subject does not increase accuracy.

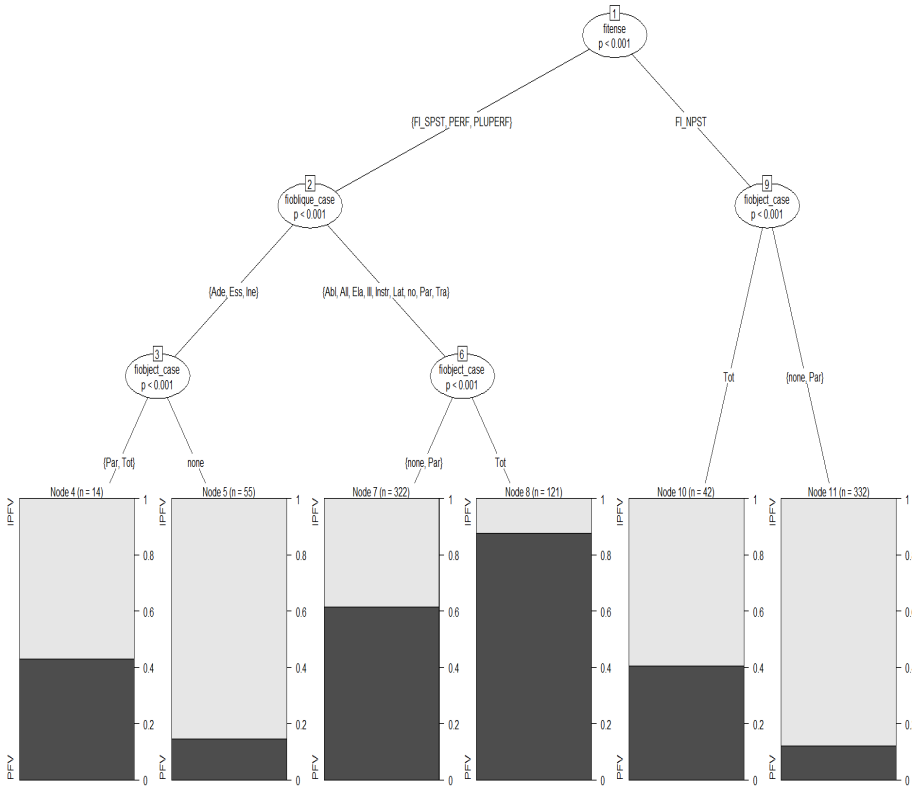

Figure 7.20: Conditional inference tree predicting PVA values exclusively from Finnish grammatical categories.

The tree divides the data first according to tense, either the Non-past and past tenses (Node 1). This corroborates the results of the cluster analysis concerning the correlation between aspect and tense. Node 2 defines the rule for splitting clauses formulated in Finnish past tenses. Since one branch contains Adessive, Essive and Inessive (essive cases), and the other Ablative, Allative, Elative, Illative, Lative, Translative (lative cases), I can conclude that the factor important for the split in Node 2 is essive-lative opposition, held by Kangasmaa-Minn (1984) and Tommola (1986) as important for determining the aspect of a clause in Finnish. It is important to observe that both state cases are included in this node, which speaks in favour of treating the Translative and Essive as spatial cases, as recommended by Siro (1964) and Tommola (1986). In the group of past-tense clauses with essive arguments, the important distinction on Node 3 is made between clauses with nominal object, no object, or an embedded object clause. Clauses in past tenses with essive argument are generally interpreted in Polish as IPFV, but some observations have corresponding PFV clauses. The share of such PFV clauses is higher for clauses with a nominal object (Terminal Node 4), but this node is also quite small (n=14), while Terminal Node 5 contains 55 observations.

Past-tense clauses without essive argument are mostly PFV. The probability that such a Finnish clause has an aligned Polish clause containing PFV is higher for clauses with Total object (Terminal Node 8). Thus DOM, which is used as a predictor in Node 6, is significant for past-tense clauses which do not contain an essive argument.

Oblique-argument case marking is irrelevant for classification of clauses in the Finnish Non-past tense, which mainly have the parallel Polish clause with IPFV, but in the model the presence of Total object in the Finnish, Non-past tense clause increases the probability that the Polish aligned clause is PFV.

Following the recommendations of Levshina (2015: 299-300), I evaluate the goodness of fit (how well the model fits to the data) using the C-index,[16] in this case C=0.81. Values above 0.8 indicate good fit.

The model can be also evaluated by comparing the number of correctly and incorrectly classified observations. This is called ACCURACY. In the tree in Figure 7.20, its success rate is 0.76 what can be concluded from Table 7.5 by comparing the sums of diagonals in the table.

[16]Function `somers2` from R package `Hmisc` (Harrell & Dupont 2018).

	ACTUAL	
PREDICTED	IPFV	PFV
IPFV	372	71
PFV	139	304

Table 7.5: Accuracy of a single conditional inference tree in predicting PVA values from Model 1

The random forest method can be used to determine how stable the result of one tree is, that is, how much it differs from other trees that could be built. The random forest here is grown from 1000 conditional inference trees. The importance of particular predictors is shown in Figure 7.21.

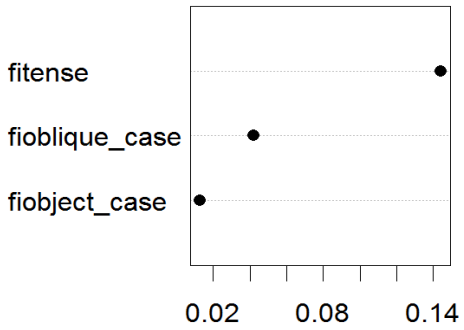

Figure 7.21: Importance of particular predictors in random forest Model 1 for predicting PVA values from Finnish grammatical categories.

Tense is the most important predictor in the random forest (variable importance 0.144),[17] while case marking of oblique arguments (0.042) and object (0.013) are less important. Thus, the results shown by the tree in Figure 7.20 are quite robust.

The random forest in Model 1 has C=0.85. The accuracy of this random forest is 0.77 (see Table 7.6). 118 IPFVs are misclassified, which is 23% of all IPFV's, and 77 PFVs (20% of all PFV). Assuming that data would be classified naively only according to the more frequent value – the IPFV – I would get 56% accu-

[17] According to what has been said about variable importance in the beginning of this section, excluding tense from the data would decrease accuracy by 0.144.

	ACTUAL	
PREDICTED	IPFV	PFV
IPFV	393	78
PFV	118	297

Table 7.6: Accuracy of random forest Model 1 in predicting PVA values

racy. Hence, applying random forest to Model 1 means an improvement of 21 percentage points in comparison to naïve classification.

7.4.3 Model 2 – Finnish grammatical and semantic features

PVA not only interacts with grammatical features, but is also associated with semantic features, especially with temporal quantification. In Model 2, I add semantic features related to the semantics of arguments and type temporal quantification. It is summarised in Table 7.7.

An example of a conditional inference tree is shown in 7.22.

Predictor	Levels	Explanation
fitense	NPAST, SPAST, PERF, PLUPERF	Finnish tenses
fiobject_case	TOT, PAR, NONE	Type of Finnish direct object
fioblique_case	ALL, ILL, ELA, ABL, INE, ADE, TRANS, ESS, INSTR, LAT, GEN, PAR, no	Case of Finnish arguments
fiobject_role	mover, patient, static	Semantic macrolabels of object
fisubject_role	agent, patient, positioner, receiver, mover, static	Semantic macrolabels of subject
fi_oblique_role	agent, beneficiary, circumstance, cogniser, container, content, stimulus, direction, degree, location, patient, instrument, manner, path, perceiver, recipient, reference, source, topic	Semantic labels of oblique argument
temp_quant	specific, non-specific, statement, pattern	Types of temporal quantification

Table 7.7: Predictors and their levels in Model 2

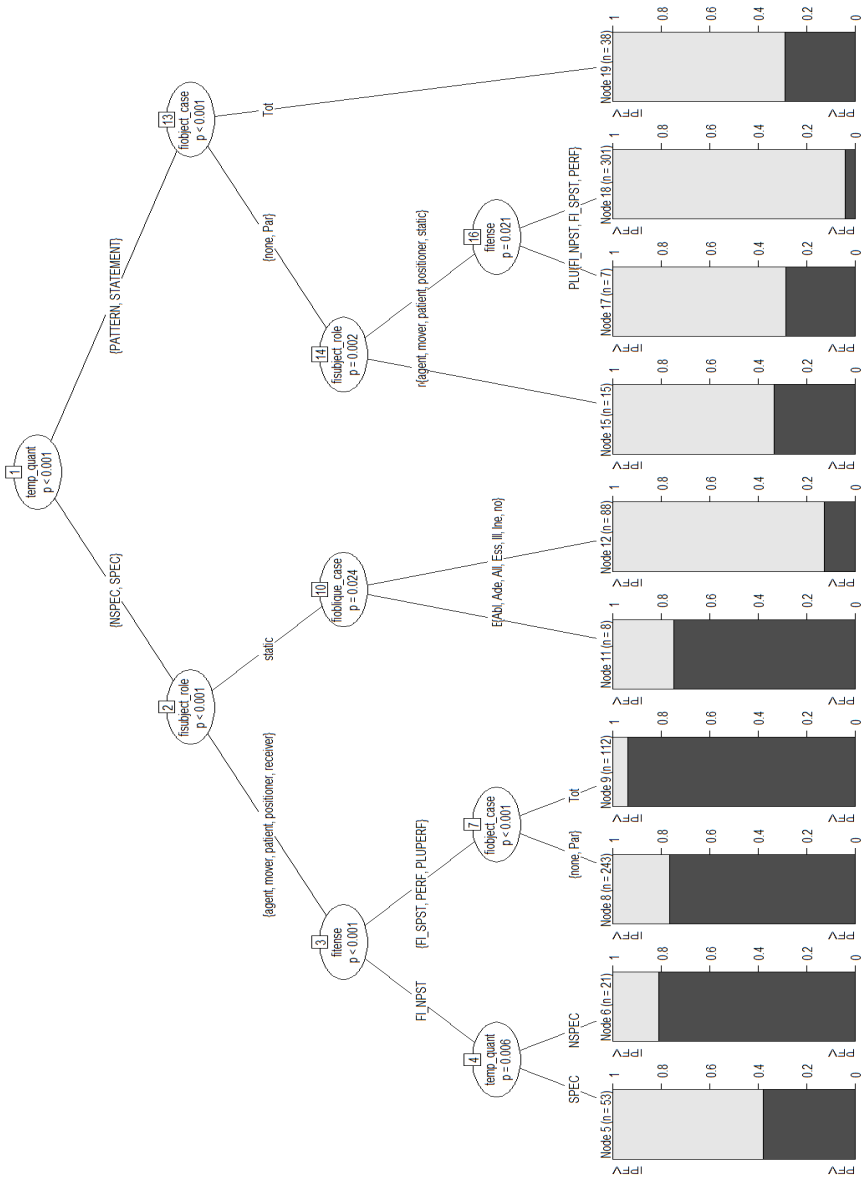

Figure 7.22: Conditional inference tree predicting PVA values from Finnish grammatical categories and semantic information.

The split on Node 1 is according to the main types of temporal quantification (specific and non-specific against patterns and statements). Yet, according to this split, I cannot generalise that Finnish existentially quantified clauses correspond with the Polish clauses in IPFV, because Nodes 5 and 12 contain mostly Finnish clauses aligned with IPFV. Therefore, the second split on Node 2 is according to the semantic macrolabel of the subject. A static subject (see Appendix G) leads to Node 10, while agent, patient, positioner, receiver, and mover are directed to Node 3. Therefore, the rule for this node uses the opposition dynamic-static as characteristic of subject in the clause. The dynamic-static opposition (Section 3.6) is broadly used to characterise situations (Vendler 1957; Hakulinen et al. 2004: §1501) or the semantics of verbal lexemes (Laskowski 1998b: 156). Here, it is visible that this opposition is aspectually relevant, and encoded in the semantic role of the subject. Existentially quantified clauses with a static subject are generally aligned with IPFV Polish clauses. Node 10 contains 96 observations while Node 11 contains only observations with the Elative oblique argument which are more likely to be aligned with PFV. Although this is based on a very small number of observations (Terminal Node 11 contains only 8) the inspection of clauses confirms that this node is perfectly interpretable in terms of aspect. The lexeme *tulla* 'come', found in 7 clauses, is used in the Finnish construction ELA + *tulla* + NOM/PAR in which the Elative argument is the patient which undergoes a change characterised by the end state in argument in the Nominative. In the current study this was treated as the subject and assigned to the semantic label reference. An example of such a clause is given below in (120).

(120) a. *Zostaniesz* *moją* *wspólniczką.*
with.PFV.stand.~~IPFV~~.2SG partner.INS
b. *Sinusta tulee* *liittolaiseni.*
you.ELA come.3SG ally.POSS.1SG
'You will become my partner.' (S1777)

I now turn to Node 3, where clauses with a dynamic subject are divided further according to tense. Non-past clauses are split again with regard to the temporal quantification (Node 4). Finnish clauses with non-specific temporal quantification are more likely to have a clause aligned with PFV, while clauses with specific existential quantification are more likely in the Finnish Non-past tense to be aligned to the Polish clause with IPFV. This split could be explained as follows. In Polish, PFV existentially quantified clauses with the Non-past tense refer to situations in

the future. Since future is unknown to the speaker, he is less likely to make statements about particular moments in time. However, when a clause uses IPFV it is most probably assigning a TSIT to a TU. TU is the parameter always known to the speaker, so the clause must have a specific temporal quantification.

Having analysed the observations quantified existentially, I now examine Node 13 containing clauses quantified universally. Such Finnish clauses are mostly aligned with Polish clauses containing IPFV (see Section 7.3.4). Most of them (n=301) end up in terminal Node 18 where Finnish clauses are aligned to the Polish clauses with IPFV. Nonetheless, three small groups can be distinguished where the probability of PFV is significantly higher than in Node 18. First, on Node 13, DOM is again used as significant predictor. Clauses with the Total object (Terminal Node 19, n=38) are more likely to have IPFV Polish counterparts. Clauses without the Total object are divided on Node 14 into clauses with a recipient subject (Terminal Node 15, n=15) and other subjects. Hence, the presence of recipient subject increases the chances that the Finnish clause is aligned in the data to a Polish clause with IPFV. Finally, Node 16 draws a distinction between the Pluperfect and other tenses. Although Finnish clauses aligned to Polish clauses with PFV appear here too, this node is very small (n=7).

The tree fits the data well (C=0.90). Table 7.8 shows the accuracy of the example tree.

	ACTUAL	
PREDICTED	IPFV	PFV
IPFV	441	61
PFV	70	314

Table 7.8: Accuracy of a single conditional inference tree in predicting PVA values from Model 2

Adding the type of temporal quantification and semantic macrolabels of the subject, object and the semantic labels of oblique arguments increases the accuracy of the model to 86%, while the importance of tense drops.[18] Two semantic predictors are relevant for the improvement: temporal quantification and the semantic macrolabels of the subject. Other semantic labels are not used, which is

[18]I tried several variables of Model 2. Abandoning tense and temporal quantification decreases the accuracy to 74%, while leaving out only temporal quantification brings back the importance of tense for the model, with 81% accuracy. Random forest slightly simplified by reducing temporal quantification to the difference between existential and universal performs only slightly worse than the reported model with 85% accuracy.

in accordance with the statements made in Section 7.3.3. The semantic labels of object and oblique arguments are partially redundant with semantic labels of subject. In other words, the label of the subject limits the scope of allowed labels of other arguments in the clause. Secondly, semantic labels of oblique arguments are partially redundant with the case marking of oblique arguments.

The variable importance for the random forest (again 1000 trees) shown in Figure 7.23 confirms that a single tree is quite robust. After temporal quantification (variable importance 0.031), the second most important predictor is semantics of subject (0.022) which is rated over tense (0.015). Object case is the next most important predictor (0.004), while the role of object and oblique case have relatively little impact (0.001). The semantic labels of oblique arguments are not significant for the model.

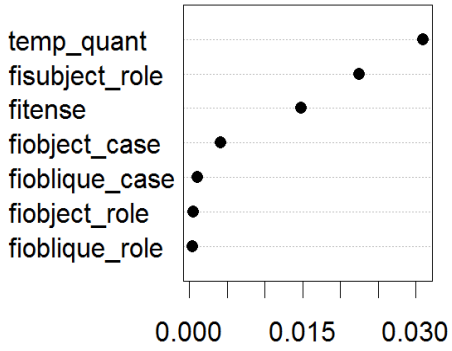

Figure 7.23: Variable importance for a random forest built upon Model 2 predicting PVA from Finnish grammatical categories and semantic information.

The random forest built in Model 2 has C=0.94, and it predicts with 87% accuracy (Table 7.9). In all, 65 IPFVs (12.7%), as are 55 PFVs (14%) are misclassified. Thus, adding semantic information improves the overall model, but it contributes more to the accurate classification of IPFV, where the error rate is now nearly 10 percentage points lower, while PFV improves only by 6 percentage points.

	ACTUAL	
PREDICTED	IPFV	PFV
IPFV	446	54
PFV	65	321

Table 7.9: Accuracy in predicting PVA values of a random forest built on Model 2

This leads to the conclusion that PFV has better grammatical correlates in Finnish than IPFV. PFV can be better predicted exclusively from grammatical features, and extending the model of semantic variables improves PFV prediction less than IPFV prediction.

7.4.4 PVA and theme-to-event homomorphism

The study of the parallel corpus shows the correlation between PFV and TOT. In the previous models, the case marking and the role of the object were not the most important factors, due to the interaction between tense, aspect, and temporal quantification. Additionally, DOM may be used as a predictor only in transitive clauses. Similar models could thus be built only for transitive clauses. Let me examine the variable importance in random forests built on the same set of predictors as Model 1 but for transitive clauses (left plot in Figure 7.24), and for transitive clauses where theme-to-event homomorphism should be mostly possible (right plot in Figure 7.24), that is, for objects with the semantic macroroles patient and mover.

Comparing the results from Figure 7.24 to those from Figure 7.21, the importance of DOM (0.5) in the prediction clearly increases in transitive clauses above oblique-argument case marking (0.001). The variable importance of DOM is even higher in the clauses which allow for theme-to-event homomorphism (0.073), whereas oblique case has a negative value (-0.003), which means that in geneal adding that variable to the forest decreases the accuracy of prediction. Although the importance of tense remains higher in the hierarchy in all three models, in clauses with typical theme-to-event homomorphism, type of object is very close to tense, which has now only 0.088 variable importance (it was 0.144 in the model with all observations, and 0.115 in transitive clauses).

Both forests have a similar goodness of fit in predicting PVA; for transitive clauses C=0.86, for theme-to-event homomorphic clauses C=0.87. Accuracy in

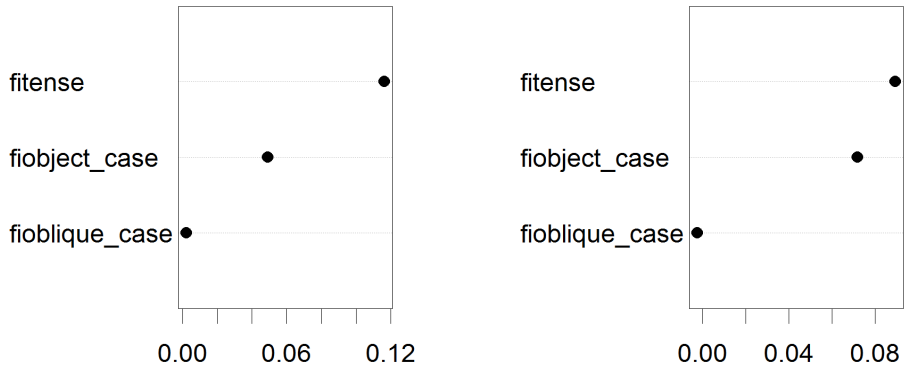

Figure 7.24: Variable importance for a random forest predicting PVA values from Finnish grammatical features in transitive clauses (left) and transitive clauses with typical semantic labels allowing for theme-to-event homomorphism (right).

both models reaches 0.78 (Tables 7.10 and 7.11). Thus models for these two datasets perform marginally better than the general model for all kinds of clauses.

	ACTUAL	
PREDICTED	IPFV	PFV
IPFV	99	29
PFV	43	150

Table 7.10: Accuracy in predicting PVA values of a random forest for transitive clauses

The error rate yields some insights (Table 7.10). A random forest built for all transitive clauses is a good classifier for PFV (29 times misclassified, 16%), but performs very poorly for IPFVs, where 29% (43 observations) are misclassified. Thus, the proximity between PFV and TOT is visible in this random forest.

The random forest for semantic macrolabels which clearly allow for theme-to-event homomorphism behaves the opposite way to the random forest for all transitive clauses (Table 7.11). IPFV is predicted better (5 errors, 12%) than PFV (17 errors, 28%). This result can be explained by the fact that the overall number of observations here is only 101; for instance, not all tense-object combinations are equally well represented. Nonetheless, it is clear that DOM is a relevant predictor of PVA and can be treated as its grammatical correlate.

	ACTUAL	
PREDICTED	IPFV	PFV
IPFV	36	17
PFV	5	43

Table 7.11: Accuracy in predicting PVA values of a random forest built for clauses which should allow theme-to-event homomorphism

7.5 Conclusions

Based on this observations, it is now possible answer to the empirical research questions formulated in the beginning of this chapter.

A1 As long as an indicative, affirmative clause is considered, the best quantitative correlates of PVA in Finnish are tenses and argument cases. Finnish tenses are stronger correlates than argument cases. In particular, the Finnish Non-past tense correlates with IPFV, while the Finnish Simple Past correlates with PFV. Finnish argument cases are weaker correlates than tense, that is, tense explains more variance and is in general a better predictor of PVA value. The rules of correlation for the Finnish oblique cases are not necessarily very stable, but essive cases seem to correlate with IPFV, while lative cases correlate with PFV.

A2 Finnish DOM is a phenomenon belonging to the argument case marking. I identify a correlation between the Total object and PFV. Thus, in transitive clauses, DOM allows for theme-to-event homomorphism. Nevertheless, DOM is less important than tense in determining aspectual values. First, due to the interaction between tense,aspect, and temporal quantification. Second, due to the frequency. While tense is an obligatory category, DOM is not.

A3&4 Since accounting for the combination of tense *and* argument cases gives significantly better results, it can be concluded that PVA is correlated with a cluster of features. However, PVA is similarly closely correlated with Polish tenses. herefore, in indicative clauses with one finite verbal form a cluster-to-cluster correlation takes place. It comprises of Polish tense-aspect forms, as one cluster, correlating with tense and argument case-marking in Finnish, as the second.

A5 The closest semantic correlate of PVA identified in the cluster analysis is temporal quantification. Specific temporal quantification correlates with PFV value, whereas the static character of subject and patterns (subtype of universal temporal quantification) correlates with IPFV value.

However, the correlation between PVA and temporal quantification is not a two-way interaction, but a three-way interaction which requires accounting for

tense. This is because PFV correlates with specific temporal quantification only in the Polish Past tense. Also the correlation between IPFV and universal quantification is stronger in the Non-past tense than in the Polish Past tense.

Chapter 8

Conclusions and future research possibilities

8.1 Introduction

The aim of this final chapter is to evaluate the results of the quantitative study in the context of the research goals formulated in Section 1.2. Therefore, I consider how the answers to the empirical research questions (see Sections 7.1 and 7.5) contribute to the overall knowledge about PVA as a language-specific category (Section 8.2), and as a concept of which Finnish equivalents are studied (Section 8.3). Then, I revise the semantics of PVA in terms of formal marking (Section 8.4) and scalarity (Section 8.5). I comment on whether DOM could be described in terms of scale. Finally, I evaluate how the present work helps in understanding aspectuality as a semantic, cross-linguistically valid concept (Section 8.6).

For various reasons, not all problems could be examined as deeply as I would have wished. Therefore, I summarise the limitations of the study in Section 8.7 and suggest some topics for further studies in Section 8.8.

8.2 PVA

8.2.1 PVA as a multi-layered category

The study confirms the validity of the multi-layered concepts of PVA noted by Holvoet (1989): the temporal-deictic component and the inner component related to the predicate-argument structure should be distinguished. I discuss each component below.

8.2.2 The temporal-deictic component of PVA

The legitimacy of distinguishing the temporal-deictic meaning in the context of PVA has been confirmed in the hierarchical cluster analysis of distance matrix (see Section 7.3.4) where tense turned out to be a very close correlate, next to the aspectual forms (prefixes and simplex verbs).

Secondly, I identified the correlation between PFV and existential temporal quantification (subdomain of pluractionality) which does not seem to have any other separate, grammatical marker in Polish. PFV is more frequently used to refer to situations happening in time (existential quantification) which have a past temporal reference expressed in the Past tense form, while IPFV is used for situations continuously valid in time (types, universal quantification). This is therefore a three-way interaction which involves PVA, tense and temporal quantification.

8.2.3 The inner component of PVA

In the domain of predicate-argument structure, the semantics of subject, which in the current model was considered an obligatory syntactic element of clause, correlates with PVA. In the dendrogram (Figure 7.19), static subjects, which do not cause or undergo change, are placed closely to IPFV.

A pattern not retrievable from the dendrogram, but nevertheless present in the distance matrix, concerns the relation between PFV and arguments. All arguments included in the model, apart from static subjects, are closer to PFV than to IPFV. This relation is the strongest between PFV and the Finnish Total object which meaning was characterised (Section 4.3.3) as reaching the "perfect state" in situation, or material bound (Lindstedt 1995). This result suggests that the existence of a material bound is relevant for the meaning of PFV.

Hence, on the level of predicate-argument structure, PVA is characterised by the grade of change in time, from static (no change in time), through dynamic (change observable but the material bound not reached), to change observable (material bound reached).

8.3 The correlates of PVA in Finnish

8.3.1 Tense as correlate

Having specified the semantics of PVA, it is now easier to compare Finnish and Polish following the same logic of layers. On the temporal-deictic layer, Finnish

tenses are clearly the closest counterparts of PVA.

Neither Polish nor Finnish can benefit from a system of articles or classifiers which could contribute to marking temporal quantification. Among the language tools of Finnish, tense is the closest statistical correlate of PVA. Although Finnish makes use of four tenses, the Non-past and the Simple Past are much more frequent than the Perfect and the Pluperfect. The distributions of the former two tenses in corpus_2 are very similar to the distributions of Polish tenses, so similar patterns can be identified to those found in Polish, but due to the lack of aspectual verbal marker, the interaction only goes two ways. Existential quantification dominates in the Simple Past, while universal quantification dominates in the Non-past tense. This explains why PFV is close to the Finnish Simple Past, while IPFV is close to the Finnish Non-past.

8.3.2 Elements of the Finnish predicate-argument structure

The two semantic parameters now identified as relevant for the predicate- argument structure and PVA are the notion of change and the grade of change in time. As to predicting the value of PVA in the clause, I identified that, within the predicate-argument structure, case marking plays a significant role: on the one hand in DOM, and on the other, in lative-essive opposition. Both notions have been held as aspectually relevant by other scholars (Dahl 1985; Heinämäki 1994; Kangasmaa-Minn 1984; Tommola 1986; Zmrzlíková 2009).

Hence, the static component characteristic for IPFV has a correlate in Finnish essive cases. The disproportionate number of verbal lexemes appearing in the Polish and Finnish clauses in corpus_2 (see Section 7.2.1) provides additional empirical evidence for Kangasmaa-Minn's claim as to the spatial case-marking nature of marking aspectual oppositions, and is in accordance with the work of Biskupska (2018).

The conclusions of Tommola (1986) and Zmrzlíková (2009) are not far from my results. PVA and DOM overlap on the semantic layer, but PVA has a strong temporal-deictic correlate, which DOM does not have, or which is not as strongly resembled in the frequency as it is for PVA.

8.3.3 The relation between DOM and Slavic aspect

It has been suggested that three of the Slavic languages, Czech, Russian, and Polish, differ with respect to semantic-functional domains marked by the verbal as-

pect opposition (see Section 1.2). In Section 4.3, I referred to two contrastive studies of expressing aspect in Slavic languages (Tommola 1986; Zmrzlíková 2009).

In relation to the Finnish DOM the latter two studies and the present one come to the same conclusion: the distributions of Slavic verbal aspect and DOM overlap, but they do not match entirely. All three works are based on empirical data, so it is possible to compare them briefly.[1] The percentages are provided in Table 8.1.

	CZECH		POLISH		RUSSIAN	
	TOT	PAR	TOT	PAR	TOT	PAR
PFV	72.8%	27.2%	71.3%	28.7%	72.6%	27.4%
IPFV	24.8%	75.2%	28.8%	71.1%	21.3%	79.7%

Table 8.1: Slavic verbal aspect and Finnish DOM

Tommola and Zmrzlíková obtained their data mostly from literary texts and their samples are roughly one third smaller than corpus$_2$. Unfortunately, for comparison with Russian, Tommola (1986) does not provide raw frequencies but only percentages, so no significance testing is possible and the data for PFV and IPFV must be treated as two separate populations.

However, the distributions of Czech and Polish data can be compared. The chi-square test does not allow for rejecting the hypothesis that Czech and Polish distributions are significantly different (p=0.43), so one could assume that there is no significant difference in the relationship between verbal aspects in Slavic languages and DOM in Finnish. Thus, the meaning of change reaching the material bound is most probably present in the PFV meaning of these three Slavic languages and should be acknowledged as one of the semantic components of meaning of the Slavic perfective.

In contrast to the previous work, my analysis suggests that contribution of OSMAs in expressing aspectuality is marginal, because this type of adverbial occurs rarely in language use. This might be related to the fact that the role of all lexical expressions in expressing temporal properties seems marginal in comparison to the grammatical features. Linguistic elements of this sort are relevant rather from the perspective of different discourse structures (text types) than in terms of the structure temporal system of Polish and Finnish.

[1] For Tommola (1986) I use only the real text excerpts, and not the examples taken from other scholars which were included in the original comparison, since aspect distribution does not correspond with real-language use in the excluded data.

8.4 Formal marking of PVA

8.4.1 Spatial character of change marked in Polish

The contrastive study yields findings about affixes as markers of PVA. As concluded above, Finnish expresses the static-dynamic opposition partly with semantic, spatial cases. This result is not surprising, considering that Slavic languages also mark PFV mostly spatially – with prefixes which evolved from spatial prepositions. The prefix implies the notion of change through marking the existence of some bound which prototypically was spatial, just as it is for lative cases. Since prefix is generally the carrier of the notion of change and of the bound, simplex lexemes (with exception of the close group of PFV lexemes) do not encode the notion of change related to the bound, while secondary imperfectives encode the meaning of change (concluded from the prefix) but without reaching the bound (concluded from the suffix). Thus, in the context of marking, PVA does not need to be treated as binary, but may be also reconsidered as a tripartite opposition: simplex imperfectives encode no change related to the bound, prefixed perfectives encode change reaching the bound, and secondary imperfectives encode change which does not reach the bound.

8.4.2 Temporal bound

Temporal bound (Section 4.3.2) is neiter the essence of PVA nor of Finnish case marking, but in Polish it can be realised with particular prefixes which appear in the function of delimitiative and perdurative Aktionsart (see Section 3.6), that is, the prefixes *po-*, *za-*, and *prze-*. The tools specialised in marking temporal bound in Finnish are probably measure adverbials in the object cases (Section 4.3.5), which were not included in the statistical models due to their rare occurrence in corpus_2.

8.5 Scalar approach to aspect

In Section 3.7, I used a scalar approach to characterise the PFV-IPFV opposition. In this view, PVA specifies the property of TR to which the TSIT may be assigned in the clause. PFV constrains TSIT assignment to units measured on discrete scales, so the temporal unit cannot be quantised, while IPFV deals with the temporal units measured on continuous scales.

In the light of the current study, these definitions are probably too strict, but the scope of TRs to which PFV can be assigned is certainly more restricted than the scope of TRs for IPFV. In relation to deictic temporal localising, PFV cannot be used for TRs including a TU which is expressed with the Non-past IPFV form. As to temporal quantification, PFV is preferred in existential quantification, because the TSIT of situation tokens must be assignable to unique TRs, even if the speaker is not able to point out the specific units of time.

In universal quantification, IPFV is the preferred form, because situation types are continuously valid, so if the scope of a TR includes the TU, it must be expressed in IPFV, non-past form.[2]

On the inner semantic level, the notion of change and its gradability can be described within a scalar approach. The change encoded in the situation may be measured on an interval of length one. IPFV may be used for any value of this interval, while PFV can be used for any value but zero, and its most frequent value is one.

Finnish DOM has the same mechanism, to which I referred as theme-to-event homomorphism. It measures and marks the grade of change on an object. Here, however, all values except equal to 1 are marked with the Partitive object, while the values equal to 1 are marked with the Total object.

Theme-to-event homomorphism is valid for all types of object, not only for the semantic types mover, patient and medium, where this notion is visible most vividly. Objects which I classified as static (such as stimuli of cognitive and physiological processes, e.g. loving, being hungry) which appear mostly in the Partitive object do not have a bound on the clause structure, in other words their bound is syntactically reduced (Kiparsky 2005; Tommola 1986, 1990), but it may be optionally realised in the clause in resultative constructions (see Section 4.3.4). Objects usually considered problematic (so-called *quasi-resultative* objects (Itkonen 1976), for example, of verbs *omistaa* 'to own', *sisältää* 'to contain', *huomata* 'to notice', *tietää* 'to know') always have the maximum value of the grade of change, and the Partitive is used only when the quantity of the object deviates from the maximal value.

Since the current quantitative analysis validated the relevance of DOM as correlate of PVA it is now legitimate to show the difference in the semantics of PVA and DOM in terms of scale. It lies in the way the values of the grade of change are assigned to PVA and DOM values. Assuming that x is the value of the grade of change, I could formalise the above-stated rules for Finnish DOM as follows:

[2]For patterns this means that the situation token representing the pattern can occur at any point of time, thus also at TU.

$$\{\text{OBJ} \in \text{TOT} : x = 1\} \tag{8.1}$$

$$\{\text{OBJ} \in \text{PAR} : 0 \leq x < 1\} \tag{8.2}$$

and for PVA as follows:

$$\{\text{V} \in \text{PFV} : 0 < x \leq 1\} \tag{8.3}$$

$$\{\text{V} \in \text{IPFV} : 0 \leq x \leq 1\} \tag{8.4}$$

8.6 The cross-linguistically valid domain of aspectuality

8.6.1 Multi-layered hierarchical model

Several conclusions relevant from the cross-linguistic perspective may be drawn from this study. As already discussed, aspectuality, the semantic domain, which in Polish is grammaticalised in PVA, appears to be a multi-level, hierarchical category. The two main levels concern the temporal-deictic domain, the outer, external level where temporality and aspectuality overlap, and the inner, internal domain which is closer to Comrie's (1976) definition of aspect cited in Section 1.2. In my view, a handy tool for inspecting the structure of aspectual domain is the plot of variable importance of Model 2 presented in Figure 7.23 in Section 7.4.3, which could be interpreted as hierarchical representations of semantic components of aspectuality and their language means of expression.

The main semantic components in the plot are reflected by the predictors temporal quantification, tense, role of subject, case of object and oblique. Temporal quantification and tense belong to the outer, temporal-deictic domain of aspectuality which overlaps with the semantic domain of temporality. The role of subject corresponds to the notion of change based on the opposition static versus dynamic, and finally the oblique and object argument case-marking corresponds to the grade of change.

The current map is certainly not the full map of the domain of aspectuality, as for example the notion of the temporal bound mentioned in Section 8.4.2 is peripheral in terms of expression in both Polish and Finnish but, according to Bertinetto & Delfitto (2000) and Lindstedt (1995), relevant in Romance languages. Thus, designing similar studies for various pairs of languages could improve the current hierarchical representation, and provide even broader view on the category of aspectuality as a cross-linguistically valid, comparative concept.

8.6.2 Aspectuality and the theory of oppositions

One more contribution of the current study should be pointed out here. In the previous approaches, aspectuality was based on either binary opposition (a structural approach where the semantic invariant is assigned to one marked counterpart, while the second, unmarked member is defined negatively, as lacking the particular feature) or equipollent opposition.

The current study suggests that the inner level of aspectuality can be treated as *gradual* opposition. Taking a scalar approach, I showed in Section 8.5 how gradual oppositions can be represented formally. The convenience of this solution is that it makes it possible to treat levels of aspectuality not only as binary opposition, but also as multipartite opposition, which potentially could be identified in understudied languages.

8.7 Limitations of the study

8.7.1 Corpus approach

The present corpus study was usage oriented. Thus, it does not provide any insight into acceptability, which would require other methods such as psycholinguistic experiments.

8.7.2 Size and variety

The size and variety of the studied sample that I called corpus_2 is the main obstaclepreventing some areas from being examined. The collected data covered only three text types, and did not include some features considered to be relevant from the temporal perspective. One of them was the 'Finnish progressive' construction which occurred only once. As mentioned in the previous chapter, this construction is still in use in modern Finnish, but most probably it occurs mainly in spontaneous speech. Since such language is rarely translated, parallel corpora are not an appropriate source of data. Instead, monolingual corpora could be used to identify the particular contexts when the 'Finnish progressive' appears. The construction could be then compared with PVA, by obtaining contextually comparable utterances in Polish and examining the used values of PVA.

Secondly, I limited my study to the indicative, affirmative, finite-form predicates. In this setting, the temporal (and pluractional) parameter of PVA is inseparable from tense because I dealt with situation tokens which are always con-

textualised. Consequently, conducting a comparable study on other verbal forms (e.g. imperative) could shed more light on the non-temporal functions of PVA. Such a study could probably explore better the shared semantics of PVA and the Finnish DOM.

8.7.3 Sparseness of features

Some categories such as free adv-erbals or prefixes, which comprise many items, were not sufficiently represented in the current study. Thus, it could provide no clear answer about the role of a particular item in the temporal-aspectual domain. The properties of each item (that is each prefix, each ad-verbal) would need to be examined in separate studies based on large volumes of parallel data, where each feature could occur sufficiently frequently.

8.8 Future research possibilities

8.8.1 Applying quantitative methodology to other contrastive and typological studies

The current study provided an innovative, bottom-up methodology for cross-linguistic studies of temporality and aspectuality which could be also applied in other contrastive or typological studies.

In particular, as mentioned in Chapter 1, although all Slavic languages have a category of verbal aspect, they differ in the functional-semantic scope of this category. Thus, the results of the present study could serve as a starting point for building a semantic map of aspectuality through pairwise comparisons with Finnish.

8.8.2 PVA and temporality across different text types

The data used in the present study originated from the three text types literary, informative and to-be-spoken. For a study on Slavic aspect, this material is quite broad, as most studies on Slavic aspect (if based on empirical material) restrict themselves to explaining the functions of aspect in narrative, mostly literary discourse (e.g. Tommola 1986; Zmrzlíková 2009). Aiming to draw general conclusions about the nature of PVA here, I observed some phenomena from the temporal domain which applied only to particular text types.

Firstly, in Section 5.6 I showed that PVA and tense are not equally distributed across text types. This indicates that the two grammatical categories form temporal frames in discourse structures.

On the microlevel of a particular text type, some fine-grained notions are visible. Semelfactive and momentanous morphemes could be identified only in the literary text type. The increased frequency of durative temporal expressions also characterised this subsample. Informative texts, in contrast, contained more temporal localising expressions.

In TBS texts, deictic lexical markers of temporal localising, the Finnish Perfect in the context of non-specific quantification, and the Polish verbal prefix *z-* were used more frequently than in the other two samples.

These distributions show that temporal markers, among them aspect, are not equally represented across text types, but each type prefers some particular set of temporal tools to express the relevant temporal categories. In that context PVA may serve different functions in different text types. Thus, the studies of PVA in text types other than narrative are needed, as stated in Section 8.7, in particular in oral, interactional discourse.

8.8.3 Law of interference in temporal domain

In this study, I avoided the potential impact of the law of interference (see Section 5.2.2) by balancing the directionality of parallel texts in corpus_2. Appendix C merely touches upon the possible effects in the data, and surprisingly, does not show any deviations as to the distributions of PVA and DOM. Taking into account previous research, these results are surprising. The topic certainly requires further investigations. It is possible that Toury's law of interference mostly affects very specialised linguistic units, whereas its impact in the domain of the most fundamental grammatical categories of the highest frequency is minimal.

8.8.4 Word order and information structure

For the sake of simplicity, I eliminated word order from the current study. Nonetheless, moving an object or oblique argument to the first position in a sentence changes the information structure, and can be a relevant factor influencing the interpretation of temporal quantification in a sentence (Vilkuna 1992: 179-185). Some observations concerning the role of locative expressions in the semantics of sentences and word order has been described by Huumo (1997).

Since PVA appears to be a relevant factor in the domain of temporal quantification, word order could be added as a relevant factor in contrastive studies on aspectuality in Finnish.

8.8.5 Aspect as potential challenge for natural language processing and machine translation

The present study has shown that, regardless of its temporal nature, the semantics of the grammatical category PVA is a quite abstract concept and it does not have any clear quantitative correlates in Finnish. Due to the above, I conclude that aspectuality belongs to concepts causing errors in natural language processing (e.g. machine translation) and could be a relevant topic for machine learning algorithms.

Appendix A

Main derivative verbal classes in Finnish

Typical suffixes	Derivative Class	Derivative base	Class meaning	Example
detransitivisers				
	reflexive	V	an agent is the target	*pukeutu-* 'to dress oneself'
	passive-reflexive	V	hidden, or non-existing agent	*avautu-* 'to open'
	decausative	V		
-ne-, -UtU-, -VntU-, -OitU-, -ntU-, -(i)ttU-, -(i)stU-, -(t)U-	translative	ADJ	to gain a property marked by ADJ	*suurene-* 'to become big(ger)' from *suuri* ('big')
-VntU-, -UtU, -OitU-, -(i)ttU-, -(i)stU-, -(t)U-	receptive	N	being under the influence of a quality described in the root	*velkaantu-* 'run into debt'
-UtU-, -VntU-, -ne-	directional	N	aimed location	*rantautu-* 'go / come ashore'
transitivisers				
-ttA-, -UttA-, -tUttA-	causative	V		*syötä-* 'to feed',

-tta-, -UttA-, -tUttA-, -ltA-, -stA-, -	curative-causative	V/N	emotions, desideratives, Partitive-experiencer	*janotta-* ‘make sb thirsty’, *ärsyttä-* ‘make sb angry’
	resultative	N/ADJ	describes the aimed result included in the root	*nopeutta-* ‘to accelerate’ from *nopea* ‘fast’), *hankaloitta-* ‘to complicate’ from *hankala* ‘inconvenient, troublesome’
	purposive	N	to supply with a thing or phenomenon described by the root	*normitta-* ‘to norm’
	privative	N	erasing something	*pölyttä-* ‘to dedust’ from *pöly* ‘dust’
	instrumental	N	to act with a tool described in the base	*puukotta-* ‘to stab’ from *puukko* ‘knife’
	directional	N	aimed location	*nauhoitta-* ‘to record’ from *nauha* ‘tape’, *piilotta-* ‘to hide’ from *piilo* ‘hiding place’
-ksU-	attitude	ADJ	to consider as something	*paheksu-* ‘to disprove’ from *paha* ‘bad’
temporal structure modifiers				

-ele-, -ile-, -entele-, -skele-, -skentele, -i-, -o-, -ise-	frequentative	V	repeated action	*hyppele-* ‘to hop about’
-AhtA, -Aise-	momentaneous	V	short-lasting, sudden, rapid actions	*käännähtä-* ‘to turn suddenly’,
-i-, -o-, -ksi-	continuative	V	continuing situations	*hypi-* ‘to jump for a while’
other				
-Oi-, -le-, -tA-	essives	N/ADJ		*sairasta-* ‘to be ill’

Table A.1: Main derivative verbal classes in Finnish based on Koivisto (2013: 289) and Kiefer & Laakso (2014); Laakso (1997)

Appendix B

Texts included in corpus$_2$

S1–S108			
Original	Text type		INFORMATIVE
	Language		FINNISH
	Number clauses	of	83
	Author		Lehtonen, Juhani
	Title		*Suomen kansallistunne*
	Year/date		2005
	Published in/at		Hiidenkivi: suomen kulttuurilehti 12(6), 6–8
Translation	Translator		Gąsiorowska-Siudzińska, Mariola
	Title		*Cechy narodowe Finów*
	Year/date		2008
	Published in/at		Czas Kultury : kultura literatura filozofia 24 (2), 10–15.
S109–226			
Original	Text type		INFORMATIVE
	Language		PL
	Number clauses	of	93
	Author		Kapuściński, Ryszard
	Title		*Tłumacz - postać XXI wieku*
	Year/date		16.12.2013
	Published in/at		`http://serwisy.gazeta.pl/kapuscinski/1\&2308/4\&2746460.html?as=2Ã§ixzz2nedP74Zc`

Translation	Translator	Kärkkäinen, Tapani
	Title	*Kääntäjä 2000-luvulla*
	Year/date	16.12.2013
	Published in/at	`http://www.saunalahti.fi/tapank/index2.html`
S253–S281		
Original	Text type	INFORMATIVE
	Language	FINNISH
	Number of clauses	25
	Author	unknown
	Title	*Suurlähetystön näyttely esittelee Mannerheimin onnelliset vuodet Puolassa*
	Year/date	05.06.2013
	Published in/at	`http://www.finland.pl/public/default.aspx?contentid=277908\&nodeid=40900\&contentlan=1\&culture=fi-FI`
Translation	Translator	unknown
	Title	*Wystawa w Ambasadzie prezentuje Marszałka Mannerheima podczas szczęśliwych lat spędzonych w Polsce*
	Year/date	12.06.2013
	Published in/at	`http://www.finland.pl/public/default.aspx?contentid=278454\&nodeid=40901\&contentlan=22\&culture=pl-PL`
S335–S344		
Original	Text type	INFORMATIVE
	Language	FINNISH
	Number of clauses	7
	Author	unknown
	Title	*Suurlähettilään maakuntamatkat jatkuvat*
	Year/date	22.03.2013

	Published in/at	http://www.finland.pl/public/default.aspx?contentid=273111\&nodeid=40900\&contentlan=1\&culture=fi-FI
Translation	Translator	unknown
	Title	*Ambasador Vilen z wizytą w Uniejowie i Łodzi*
	Year/date	21.03.2013
	Published in/at	http://www.finland.pl/public/default.aspx?contentid=272957\&nodeid=40900\&contentlan=22\&culture=pl-PL
S346–S355		
Original	Text type	INFORMATIVE
	Language	FINNISH
	Number of clauses	10
	Author	unknown
	Title	*Summit 2010*
	Year/date	03.07.2013
	Published in/at	http://www.finland.pl/public/default.aspx?nodeid=42375\&contentlan=1\&culture=fi-FI
Translation	Translator	unknown
	Title	*Summit 2010*
	Year/date	03.07.2013
	Published in/at	http://www.finland.pl/public/default.aspx?nodeid=42375\&contentlan=22\&culture=pl-PL
S365–S375		
Original	Text type	INFORMATIVE
	Language	FINNISH
	Number of clauses	9
	Author	Sivonen, Jari
	Title	*Suurlähettiään terveys*
	Year/date	03.07.2013

	Published in/at	http://www.finland.pl/public/default.aspx?nodeid=40906\&contentlan=1\&culture=fi-FI
Translation	Translator	unknown
	Title	*Powitanie ambasadora*
	Year/date	03.07.2013
	Published in/at	http://www.finland.pl/public/default.aspx?nodeid=40906\&contentlan=22\&culture=pl-PL
S379–S399		
Original	Text type	INFORMATIVE
	Language	FINNISH
	Number of clauses	16
	Author	unknown
	Title	*Puolan vesinäytteiden tulokset viittaavat merkittäviin fosforipäästöihin kipsivuorista*
	Year/date	01.09.2013
	Published in/at	http://www.finland.pl/public/default.aspx?contentid=280424\&nodeid=40900\&contentlan=1\&culture=fi-FI
Translation	Translator	unknown
	Title	*Wyniki próbek wody pobranych w Polsce wskazują na znaczną ilość fosforu wyciekającego z hałd fosfogipsu*
	Year/date	01.09.2013
	Published in/at	http://www.finland.pl/public/default.aspx?contentid=280448\&nodeid=40901\&contentlan=22\&culture=pl-PL
S820–S823		
Original	Text type	INFORMATIVE
	Language	POLISH
	Number of clauses	4
	Author	unknown
	Title	*Kolekcja pająków z Polski w maretarium w Kotka*

	Year/date	20.09.2013
	Published in/at	http://finlandia.msz.gov.pl/pl/aktualnosci/kolekcja_pajakow_z_polski__w_maretarium_w__kotka
Translation	Translator	unknown
	Title	*Hämähäkkejä Puolasta Kotkan maretariumissa*
	Year/date	20.09.2013
	Published in/at	http://www.helsinki.msz.gov.pl/fi/ajankohtaiset_asiat/hamahakkeja_puolasta_kotkan_maretariumissa_
S844–S848		
Original	Text type	INFORMATIVE
	Language	POLISH
	Number of clauses	4
	Author	unknown
	Title	*Targi*
	Year/date	20.09.2013
	Published in/at	http://www.helsinki.msz.gov.pl/fi/ajankohtaiset_asiat/targi
Translation	Translator	unknown
	Title	*Matkamessut*
	Year/date	20.09.2013
	Published in/at	http://finlandia.msz.gov.pl/pl/aktualnosci/matka
S849–S860		
Original	Text type	INFORMATIVE
	Language	POLISH
	Number of clauses	12
	Author	unknown
	Title	*Napięty dzień minister Kiuru w zaśnieżonej Warszawie*
	Year/date	20.09.2013

	Published in/at	http://finlandia.msz.gov.pl/pl/aktualnosci/napiety_dzien_minister_kiuru_w_zasniezonej_warszawie
Translation	Translator	unknown
	Title	*Ministeri Kiurun kiireinen päivä lumisessa Varsovassa*
	Year/date	20.09.2013
	Published in/at	http://www.helsinki.msz.gov.pl/fi/ajankohtaiset_asiat/ministeri_kiurun_kiireinen_paiva_lumisessa_varsovassa
S862–S881		
Original	Text type	INFORMATIVE
	Language	POLISH
	Number of clauses	17
	Author	unknown
	Title	*O ambasadzie*
	Year/date	20.09.2013
	Published in/at	http://finlandia.msz.gov.pl/pl/o_ambasadzie/
Translation	Translator	unknown
	Title	*Suurlähetystö*
	Year/date	20.09.2013
	Published in/at	http://www.helsinki.msz.gov.pl/fi/suurlahetysto/
S884–S983		
Original	Text type	TBS
	Language	FINNISH
	Number of clauses	98
	Author	Lehikoinen, Kirsi
	Title	*Mies vailla menneisyyttä*
	Year/date	
	Published in/at	opensubtitles.org
Translation	Translator	Gregorowicz, Katarzyna
	Title	*Człowiek bez przeszłości*

	Year/date	
	Published in/at	`opensubtitles.org`
S1028–S1039		
Original	Text type	INFORMATIVE
	Language	POLISH
	Number of clauses	12
	Author	unknown
	Title	*Historia*
	Year/date	25.09.2013
	Published in/at	`http://www.tme.eu/pl/pages/AboutCompany:1.html`
Translation	Translator	unknown
	Title	*Historia*
	Year/date	25.09.2013
	Published in/at	`http://www.tme.eu/fi/pages/AboutCompany:1.html`
S1219–S1226		
Original	Text type	INFORMATIVE
	Language	POLISH
	Number of clauses	8
	Author	unknown
	Title	*TME laureatem rankingu Dobra Firma!*
	Year/date	25.09.2013
	Published in/at	`http://www.tme.eu/pl/pages/tme-laureatem-rankingu-dobra-firma\_.html`
Translation	Translator	unknown
	Title	*TME voitti Hyvä Yhtiö rankingin*
	Year/date	25.09.2013
	Published in/at	`http://www.tme.eu/fi/pages/tme-voitti-hyva-yhtio-rankingin\_.html`
S1233–S1304		
Original	Text type	LITERARY

	Language	POLISH
	Number of clauses	59
	Author	Gretkowska, Manuela
	Title	*Kabaret metafizyczny*
	Year/date	16.12.2013
	Published in/at	merlin.pl/HTML/83-87021-84-9.html
Translation	Translator	Kärkkäinen, Tapani
	Title	*Metafyysinen kabaree*
	Year/date	16.12.2013
	Published in/at	http://www.saunalahti.fi/tapank/index2.html
S1305–S1354		
Original	Text type	TBS
	Language	POLISH
	Number of clauses	48
	Author	Gretkowska, Manuela
	Title	*Kabaret metafizyczny*
	Year/date	16.12.2013
	Published in/at	merlin.pl/HTML/83-87021-84-9.html
Translation	Translator	Kärkkäinen, Tapani
	Title	*Metafyysinen kabaree*
	Year/date	16.12.2013
	Published in/at	http://www.saunalahti.fi/tapank/index2.html
S1355–S1430		
Original	Text type	LITERARY
	Language	FINNISH
	Number of clauses	50
	Author	Tontti, Jarkko
	Title	*Luokkakokous*
	Year/date	2007
	Published in/at	Helsinki: Otava 5–27
Translation	Translator	Jurkiewicz, Edyta

	Title	*Spotkanie klasowe*
	Year/date	2012
	Published in/at	Arterie 12 (1)
S1433–S1457		
Original	Text type	TBS
	Language	FINNISH
	Number of clauses	25
	Author	Tontti, Jarkko
	Title	*Luokkakokous*
	Year/date	2007
	Published in/at	Helsinki: Otava 5–27
Translation	Translator	Jurkiewicz, Edyta
	Title	*Spotkanie klasowe*
	Year/date	2012
	Published in/at	Arterie 12 (1)
S1458–S1532		
Original	Text type	LITERARY
	Language	FINNISH
	Number of clauses	62
	Author	Lehtolainen, Leena
	Title	*Luminainen*
	Year/date	1996
	Published in/at	Helsinki:Tammi, 168–173
Translation	Translator	Musielak, Sebastian
	Title	*Kobieta ze śniegu*
	Year/date	2004
	Published in/at	Gdańsk: słowo/obraz terytoria, 170–176
S1534–S1567		
Original	Text type	TBS
	Language	FINNISH
	Number of clauses	27
	Author	Lehtolainen, Leena

	Title	*Luminainen*
	Year/date	1996
	Published in/at	Helsinki:Tammi, 168–173
Translation	Translator	Musielak, Sebastian
	Title	*Kobieta ze śniegu*
	Year/date	2004
	Published in/at	Gdańsk: słowo/obraz terytoria, 170–176
S1570–S1607		
Original	Text type	LITERARY
	Language	FINNISH
	Number of clauses	38
	Author	Hirvonen, Elina
	Title	*Että hän muistaisi saman*
	Year/date	2005
	Published in/at	Helsinki: WSOY, 8–11
Translation	Translator	Kiuru, Iwona
	Title	*Przypomnij sobie*
	Year/date	2008
	Published in/at	Warszawa: Wydawnictwo W.A.B., 8–12
S1608–S1635		
Original	Text type	LITERARY
	Language	POLISH
	Number of clauses	32
	Author	Mrożek, Sławomir
	Title	Słoń: satyry
	Year/date	2007[1957]
	Published in/at	Warszawa: Noir sur Blanc, 7,8, 43–45
Translation	Translator	Kristiina Kivivuori
	Title	Elefantti:satiireja
	Year/date	1964
	Published in/at	Weilin+Göös 38-41,145–146
S1636–S1645		
Original	Text type	TBS

	Language	POLISH
	Number of clauses	8
	Author	Mrożek, Sławomir
	Title	*Słoń: satyry*
	Year/date	2007[1957]
	Published in/at	Warszawa: Noir sur Blanc, 7,8, 43–45
Translation	Translator	Kristiina Kivivuori
	Title	*Elefantti:satiireja*
	Year/date	1964
	Published in/at	Weilin+Göös 38-41,145–146
S1646–S1707		
Original	Text type	LITERARY
	Language	POLISH
	Number of clauses	59
	Author	Andrzejewski, Jerzy
	Title	*Popiół i diament*
	Year/date	1979[1948]
	Published in/at	Warszawa : Państwowy Instytut Wydawniczy, 198–200
Translation	Translator	Lahtinen, Åke
	Title	*Tuhka ja timantti*
	Year/date	1960
	Published in/at	Helsinki: Kansankulttuuri, 284–288
S1708–S1816		
Original	Text type	TBS
	Language	POLISH
	Number of clauses	94
	Author	Mrożek, Sławomir
	Title	*Tango*
	Year/date	14.10.2013[1964]
	Published in/at	`www.wolnelektury.pl`
Translation	Translator	Kärkkäinen, Tapani
	Title	*Tango*

Year/date	16.10.2013
Published in/at	private correspondence

Appendix C

Possible interference in the data

C.1 Testing for interference in the data

In Chapter 5, I mentioned the law of interference which might affect the results. The main solution used in this study is balancing the data according to the criteria of the source language. In order to find out whether any particular studied feature was truly affected, I use the analysis of contingency tables. Similarly to the analysis of temporal profiles of text types in the data, I use the chi-square test and represent Pearson residuals in mosaic plots following the same logic as in Section 5.5.

C.2 PVA

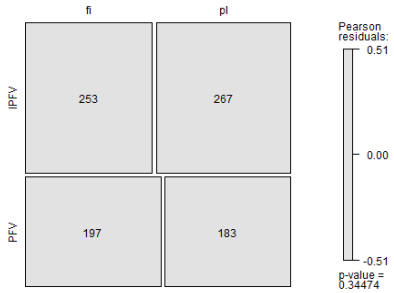

Figure C.1: Frequencies of PVA in Polish original texts (pl) and in translations from Finnish (fi).

I start this short analysis from PVA. Figure C.1 presents the distributions of PVA values across the language of original together with Pearson residuals of the chi-square test. The result does not indicate any strong deviation from expected values. Thus, the probability of interference in the domain of aspect is very small, and I can conclude that it is negligible.

C.3 Finnish derivational types, Polish derivational types, prefixes and reflexive marker

No interference can be observed on the level of Finnish and Polish verbal types, prefixes or reflexive markers, as shown below in Figures C.2, C.3 and C.4.

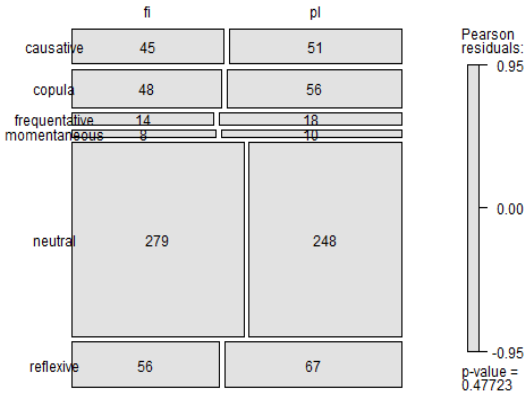

Figure C.2: Finnish derivation types in originals and translations

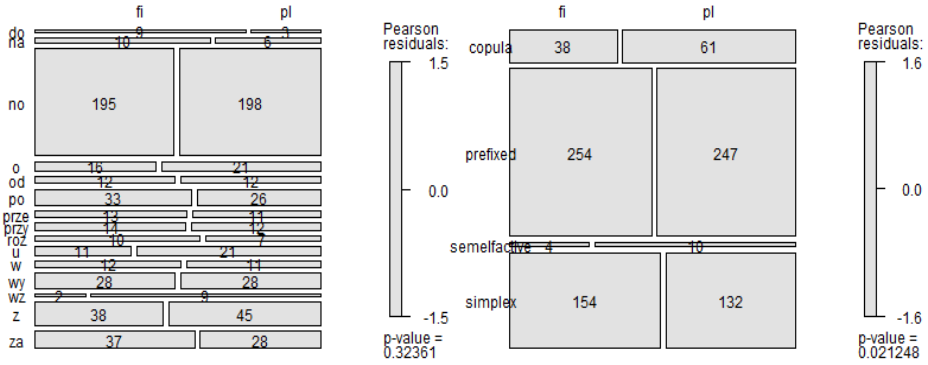

Figure C.3: Prefixes (left) and derivational types (right) in original Polish texts and in translations from Finnish

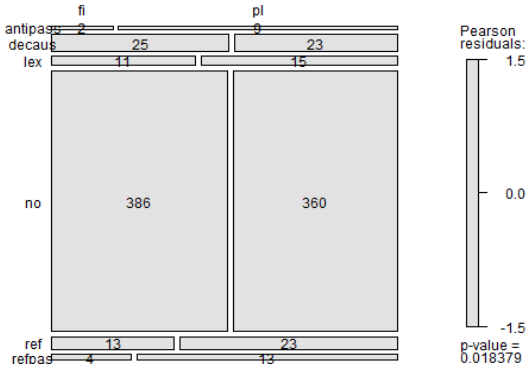

Figure C.4: Main reflexive types in original Polish texts and in translations from Finnish

C.4 Tense

The results of Polish and Finnish tense distributions across languages of originals are summarised in Figure C.5. Here, rather small differences are visible in the values of Pearson residuals.[1] First of all, the Pluperfect appears mainly in original Finnish texts, but not in translations from Polish. In translations only single sentences are written in the Pluperfect, while in Finnish originals whole

[1]The infrequent analytical future is excluded from the chi-square test

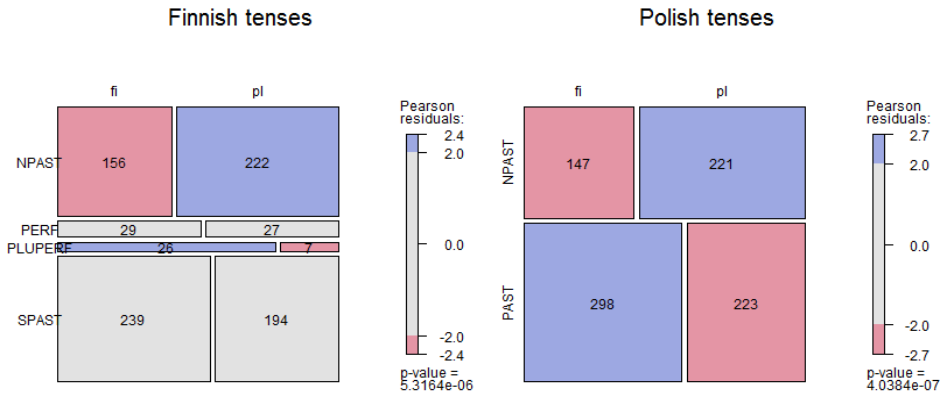

Figure C.5: Tense in originals and translations

narrative passages use this tense. Two possibilities appear: either Polish originals lacked passages where the change of narration tense is possible, or the absence of a category similar to the Pluperfect causes interference. The second reason is very probable, since also Tommola (1986: 49) observes incorrect translations from Russian to Finnish. Tommola mentions also Lingorska (1977) who points out similar problems in translation from Polish to Bulgarian.

Secondly Polish originals describe more situations with the Non-Past Form. This may be due to present temporal reference or universal quantification. Additionally, Polish has only one past tense, while Finnish has three.

C.5 Lexical expressions

As shown in Figure C.6 the frequency of the temporal localising expressions was not affected by the source language.

Localising to some absolute temporal unit is slightly more present in Finnish texts, while expressions referring to TU are more present in Polish originals.

Durative temporalisation was expressed lexically more often in Polish originals (n=14) than in Finnish originals (n=7), and at the same time such expressions were more often added in Polish translations (12 Polish expressions in comparison to 7 Finnish originals) than in Finnish translations (10 in comparison to 14 Polish originals).

The lexical expressions of single occurrence (n=3) appeared only in transla-

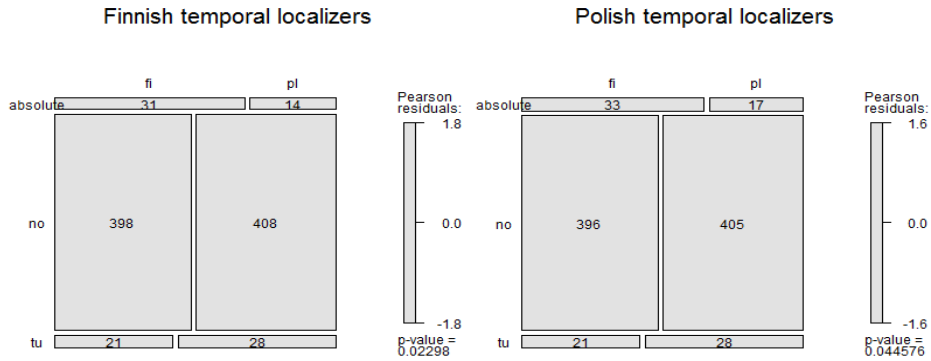

Figure C.6: Temporal localising expressions according to the source languages

tions from Finnish, which means that at least once the expression was added by the translator, because Finnish original texts contain only two such expressions in corpus$_2$. The expressions of unspecified cycle were three times more frequent in translations from Finnish (n=15) in comparison to original Polish texts (n=5), which arose from the fact that in Finnish these sort of expressions might be generally more frequent, as already described in the section above.

C.6 Taxis

The domain of taxis is not influenced by the translation as shown in Figure C.7 (next page). The frequencies of particular relations seem similar both when comparing original texts in two languages and original texts to translated texts.

C.7 Finnish object

DOM is not affected by the language of original as can read from Figure C.8.

All in all, I observe little impact of the translation process in corpus$_2$ with respect to the temporal domain. The most unbalanced results apply to the categories with few observations, thus they may be unreliable due to the small number of ob-

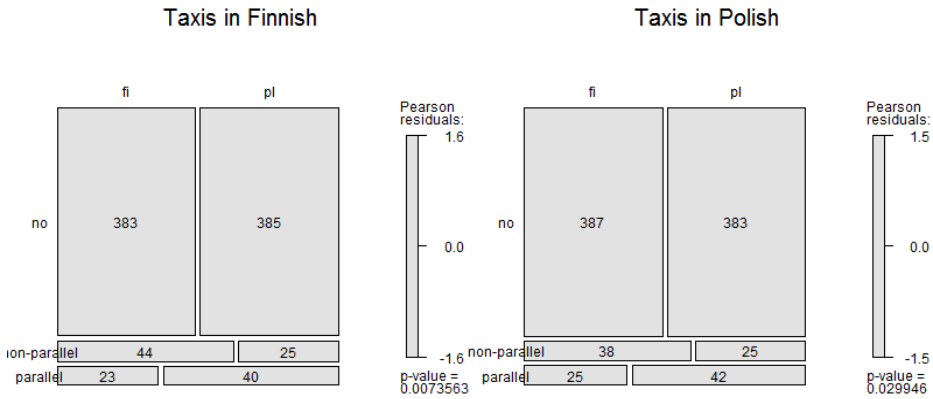

Figure C.7: Taxis according to the source languages

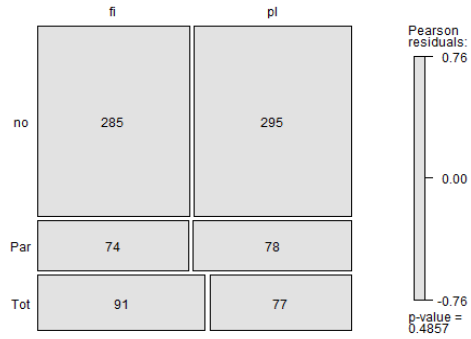

Figure C.8: The influence of original (left) and of text type (right) on DOM

servations. On the other hand, exactly these categories were not accounted for in the statistical model due to their infrequency.

My results contrast with the previously mentioned studies of Eskola (2004),McEnery et al. (2006), and Tirkkonen-Condit (2004). Although neither PVA nor DOM have straightforward counterparts, their frequencies are not affected in translation, whereas the impact on tenses, where some counterparts exist is rather small. It is possible that the effects of translation apply rather to very specialised, concrete linguistic units (particular construction or lexeme) of generally low or moderate frequency, whereas the most frequent (and therefore best acquired) grammatical

features remain unchanged.

Appendix D

A full example of a parallel clause annotated in xml

```
 <s id="1292" plstring=" Giugiu odstawił obraz na toaletkę."
fistring=" Giugiu siirsi taulun kampauspöydälle.">
    <pl>
        <plquant>specific</plquant>
        <plverb>
            <pllemma aspect="perf" derivation="prefixed">
            odstawić
            </pllemma>
            <plform>
                <simple>
                    <num>sg</num>
                    <person>ter</person>
                    <tense>praet</tense>
                </simple>
            </plform>
            <pfx>od</pfx>
        </plverb>
        <plsubj role="positioner" humscale="real"
                        spectype="ind">
            <pos>NP</pos>
            <case>nom</case>
            <num>sg</num>
        </plsubj>
        <plobj role="theme" humscale="real"
```

```
                        spectype="ind">
            <pos>NP</pos>
<case>acc</case>
<num>sg</num>
                    </plobj>
        <plarg role="direction" humscale="real"
                                spectype="ind">
            <pos>NP</pos>
            <case>na+acc</case>
            <num>sg</num>
        </plarg>
    </pl>
    <fi>
        <fiquant>specific</fiquant>
        <fiverb>
            <filemma sem_group="neutral">
            siirtää
            </filemma>
            <fiform>
                <simple>
                    <num>Sing</num>
                    <person>3</person>
                    <tense>Past</tense>
                </simple>
            </fiform>
        </fiverb>
        <fisubj role="positioner" humscale="real"
                                spectype="ind">
            <pos>NP</pos>
            <case>Nom</case>
            <num>Sing</num>
        </fisubj>
        <fiobj role="theme" humscale="real"
                            spectype="ind">
            <pos>NP</pos>
            <case>Gen</case>
            <num>Sing</num>
        </fiobj>
```

```
        <fiarg role="direction" humscale="real"
                                spectype="ind">
            <pos>NP</pos>
            <case>All</case>
            <num>Sing</num>
        </fiarg>
    </fi>
</s>
```

Appendix E

Frequency list of the Finnish verbal lexemes in corpus$_2$

filemma	frequency
copula	98
olla	84
sanoa	18
tulla	18
saada	14
tietää	14
ottaa	10
antaa	9
käydä	8
nousta	8
pitää	8
tehdä	8
jäädä	7
kertoa	7
nähdä	7
kysyä	6
mennä	6
näyttää	6
puhua	6
kulkea	5
tuntua	5
tutustua	5
viedä	5
ajatella	4
avata	4
huomata	4
kääntyä	4
katsella	4
katsoa	4
kehittyä	4
kirjoittaa	4
lukea	4
ojentaa	4
osallistua	4
pelätä	4
syntyä	4
syödä	4
tarvita	4
työntää	4
vaikuttaa	4
alkaa	3
astua	3
elää	3
heiluttaa	3
ihailla	3
istua	3
kasvaa	3
käyttää	3
keksiä	3
keskustella	3
kuolla	3
kuulla	3
kuulua	3
lähteä	3
laskea	3
löytää	3
osoittaa	3
saapua	3
selittää	3
siirtyä	3
suunnitella	3
tarjota	3
tavata	3
vahvistaa	3
vetää	3

ymmärtää	3	rakastaa	2	edetä	1
asua	2	seisoa	2	edustaa	1
esiintyä	2	sijaita	2	ehdottaa	1
esittää	2	sijoittaa	2	erehtyä	1
esittäytyä	2	soittaa	2	eristää	1
haluta	2	sopia	2	haastaa	1
heilauttaa	2	suostua	2	hakea	1
heittäytyä	2	synnyttää	2	hangata	1
huomauttaa	2	syttyä	2	harppoa	1
hymyillä	2	tappaa	2	harrastaa	1
ilahtua	2	todistaa	2	havahtua	1
ilmetä	2	toeta	2	havaita	1
jaksaa	2	toivoa	2	heijastua	1
kieltäytyä	2	tuijottaa	2	heikentää	1
korostua	2	tunnustaa	2	heilahdella	1
koskettaa	2	tuoda	2	heittää	1
kutsua	2	työllistää	2	himmentää	1
lähettää	2	työntyä	2	hipelöidä	1
loppua	2	väittää	2	hohtaa	1
loukkaantua	2	valita	2	hoitaa	1
maata	2	vierailla	2	hommata	1
maksaa	2	viettää	2	huipentua	1
merkitä	2	voittaa	2	hukkua	1
muistuttaa	2	aiheuttaa	1	huokaista	1
muuttaa	2	ajaa	1	huokata	1
myydä	2	aloittaa	1	huolehtia	1
nauttia	2	ampua	1	huudahtaa	1
nostaa	2	analysoida	1	huutaa	1
odottaa	2	änkyttää	1	hypähtää	1
päästä	2	ärähtää	1	hyppiä	1
päästää	2	ärsyttää	1	hyväksyä	1
päättää	2	arvostaa	1	ihastuttaa	1
palata	2	asettaa	1	ilmestyä	1
perustua	2	astella	1	iloita	1
piirtää	2	aueta	1	imeä	1
pystyttää	2	avartua	1	imeskellä	1
pyytää	2	diilata	1	iskeä	1
rajoittaa	2	edellyttää	1	istahtaa	1

istuutua	1	kiertää	1	liikkua	1
jäädyttää	1	kiikuttaa	1	liittoutua	1
jähmettää	1	kiinnittää	1	liittyä	1
jakaa	1	kiinnostaa	1	lisätä	1
järjestää	1	kiipeillä	1	liuottaa	1
jätättää	1	kiiruhtaa	1	loitsia	1
jatkaa	1	kiittää	1	löytyä	1
jatkua	1	kiljua	1	lumoutua	1
johdattaa	1	kimaltaa	1	luoda	1
johtaa	1	kipuilla	1	luulla	1
johtua	1	kiristyä	1	maalata	1
julistautua	1	kirjata	1	määritellä	1
julkaista	1	kirota	1	menettää	1
julkistaa	1	kivettyä	1	miehittää	1
juoda	1	kohauttaa	1	mölistä	1
juosta	1	koheltaa	1	muodostaa	1
käännellä	1	kolista	1	murahtaa	1
kääntää	1	koostua	1	murtautua	1
kaapata	1	korostaa	1	muuttua	1
kadota	1	kritisoida	1	myöntää	1
kalista	1	kruunata	1	naida	1
kamppailla	1	kuiskata	1	näkyä	1
kannustaa	1	kunnioittaa	1	napsahtaa	1
kantaa	1	kuoria	1	naurattaa	1
karata	1	kuristaa	1	nimetä	1
kärsiä	1	kurkottaa	1	nimittää	1
käsitellä	1	kuulustella	1	nipistää	1
katkaista	1	kuunnella	1	noukkia	1
katsahtaa	1	kuvata	1	noutaa	1
kävellä	1	kuvitella	1	nukkua	1
kaventaa	1	kysellä	1	nyökätä	1
käyttäytyä	1	laajentaa	1	nyökkäillä	1
keinuttaa	1	lähestyä	1	nyökytellä	1
keplotella	1	laimeta	1	ohjata	1
keskeyttää	1	läpäistä	1	oivaltaa	1
keskittyä	1	laukaista	1	opastaa	1
kieltää	1	levätä	1	osata	1
kiepauttaa	1	liikahtaa	1	osoittautua	1

ostaa	1	räpytellä	1	tunkeutua	1
otaksua	1	rauhoittaa	1	tunnistaa	1
päästellä	1	reagoida	1	tuntea	1
päättyä	1	rikastua	1	tuottaa	1
päätyä	1	rulettaa	1	turvautua	1
paeta	1	ryhtyä	1	uhata	1
paheta	1	ryssiä	1	uhrautua	1
painaa	1	saavuttaa	1	ujostella	1
painiskella	1	säilyttää	1	uskoa	1
palaa	1	säntäillä	1	vahingoittua	1
palautua	1	sataa	1	vaihtaa	1
paleltua	1	selvittää	1	vaimeta	1
palvella	1	seurata	1	vajota	1
panna	1	siirtää	1	vakuuttua	1
päteä	1	sujauttaa	1	valehdella	1
pelottaa	1	sulkea	1	valitella	1
perääntyä	1	suojella	1	vallata	1
pidentää	1	suudella	1	valmistautua	1
piillä	1	sytyttää	1	valmistua	1
pinota	1	syvetä	1	varastaa	1
pohjautua	1	taistella	1	varjostaa	1
pohtia	1	tajuta	1	vartioida	1
polkaista	1	tarttua	1	verhoutua	1
pöllyttää	1	täyttää	1	vihjata	1
pomppia	1	täyttyä	1	viipyä	1
pureskella	1	teeskennellä	1	viitata	1
purkaa	1	teollistua	1	vilkuilla	1
pysähtyä	1	toimia	1	vilkuttaa	1
pysäyttää	1	toistaa	1	vingahtaa	1
pyyhkiä	1	tönäistä	1	ympäröidä	1
raapia	1	tottua	1	yskäistä	1
rakastua	1	tukea	1		
rakentaa	1	tungeksia	1		

Appendix F

Frequency list of the Polish verbal lexemes in corpus$_2$

pllemma	frequency	podać	4	przedstawić	3
copula	99	stanowić	4	przyjść	3
mieć	32	wyglądać	4	rozwijać	3
być	15	znajdować	4	spytać	3
wiedzieć	14	dać	3	twierdzić	3
mówić	12	dostać	3	ucieszyć	3
powiedzieć	12	dostrzec	3	uczestniczyć	3
stać	9	działać	3	ulegać	3
robić	7	iść	3	uznawać	3
zostać	7	kochać	3	wskazać	3
chodzić	6	kończyć	3	wyjąć	3
wydawać	6	łączyć	3	wykazywać	3
zobaczyć	6	machnąć	3	wysuwać	3
myśleć	5	mieścić	3	zabrać	3
odwrócić	5	należeć	3	zająć	3
spojrzeć	5	nosić	3	zapytać	3
wziąć	5	omówić	3	zrodzić	3
znaleźć	5	otrzymać	3	zrozumieć	3
bać	4	podjąć	3	zwracać	3
jeść	4	potrzebować	3	żyć	3
odbyć	4	poznać	3	analizować	2
pisać	4	prowadzić	3	borykać	2

brać	2	skakać	2	delektować	1
czekać	2	skinąć	2	dilować	1
czuć	2	słyszeć	2	dobiegać	1
dochodzić	2	spotkać	2	dogadać	1
dyskutować	2	sprzedać	2	dopuścić	1
istnieć	2	stworzyć	2	dotknąć	1
jechać	2	trafić	2	dotyczyć	1
leżeć	2	trzymać	2	ewoluować	1
mieszkać	2	ucichnąć	2	gotować	1
nabierać	2	uśmiechnąć	2	grać	1
nastąpić	2	uznać	2	graniczyć	1
nazywać	2	walczyć	2	gromadzić	1
niepokoić	2	wejść	2	grozić	1
obawiać	2	westchnąć	2	irytować	1
obrazić	2	widzieć	2	izolować	1
odmówić	2	wsunąć	2	jęknąć	1
oglądać	2	wychodzić	2	kiełkować	1
ogłosić	2	wyjaśnić	2	kierować	1
ograniczać	2	wymyślić	2	kłamać	1
okazać	2	występować	2	kołysać	1
opaść	2	zabić	2	konkurować	1
opierać	2	zajmować	2	kontynuować	1
otwierać	2	założyć	2	kręcić	1
otworzyć	2	zapalić	2	królować	1
pilnować	2	zatrudniać	2	krytykować	1
podejść	2	zatrzymać	2	krzyknąć	1
pomachać	2	zwiedzić	2	kupić	1
postawić	2	błyszczeć	1	malować	1
powodować	2	boleć	1	minąć	1
przeprowadzić	2	bredzić	1	mruknąć	1
przerwać	2	bytować	1	muskać	1
przyjąć	2	całować	1	mylić	1
przyjechać	2	cenić	1	nadać	1
rosnąć	2	chcieć	1	nadgryźć	1
rozpocząć	2	cofnąć	1	nakreślić	1
rządzić	2	czytać	1	napić	1
rzucić	2	datować	1	napisać	1
schodzić	2	dawać	1	napomknąć	1

nasilić	1	piętrzyć	1	powiększyć	1
nasłuchiwać	1	płacić	1	powracać	1
następować	1	płonąć	1	powstać	1
nudzić	1	płynąć	1	powtarzać	1
obejrzeć	1	pochłaniać	1	pozostać	1
obejść	1	poczerwienieć	1	pozyskiwać	1
oberwać	1	poczęstować	1	pracować	1
oblecieć	1	poczuć	1	przebiegać	1
obronić	1	podejmować	1	przebywać	1
obsługiwać	1	podnieść	1	przechować	1
obsypać	1	podobać	1	przeczytać	1
oburzyć	1	podrapać	1	przedstawiać	1
ocknąć	1	podskakiwać	1	przejawiać	1
odbywać	1	podsunąć	1	przejść	1
oddawać	1	podziękować	1	przekazać	1
odebrać	1	podziwiać	1	przekonać	1
odnosić	1	pogłębiać	1	przekraczać	1
odnotować	1	pogorszyć	1	przenieść	1
odpalać	1	pogryzać	1	przenosić	1
odpowiadać	1	pojawić	1	przepływać	1
odrzec	1	pójść	1	przepuścić	1
odstawić	1	pokazać	1	przesłuchiwać	1
odwiedzać	1	pokiwać	1	przesunąć	1
odwiedzić	1	pokonać	1	przetłumaczyć	1
odwinąć	1	polegać	1	przetrzeć	1
odzyskać	1	pomagać	1	przewijać	1
ogarnąć	1	pomylić	1	przezwyciężyć	1
okraszać	1	popatrzeć	1	przybliżać	1
określić	1	poprosić	1	przybliżyć	1
okryć	1	porobić	1	przybyć	1
opanować	1	poruszyć	1	przychodzić	1
opublikować	1	poślubić	1	przygasić	1
osłabiać	1	postąpić	1	przygnać	1
ostać	1	postępować	1	przykładać	1
otrzymywać	1	postrzegać	1	przyłożyć	1
oznaczać	1	poświęcić	1	przynieść	1
pasować	1	potwierdzać	1	przypisywać	1
patrzyć	1	potwierdzić	1	przypominać	1

przypomnieć	1	skrzywić	1	uszczypnąć	1
przypuszczać	1	służyć	1	utonąć	1
przysiąść	1	śmieszyć	1	utrzymywać	1
przytargać	1	spać	1	utwierdzić	1
przywdziewać	1	spływać	1	utworzyć	1
przyznać	1	spodziewać	1	uwielbiać	1
przyznawać	1	spoglądać	1	uzgodnić	1
przyzwyczaić	1	spotykać	1	warknąć	1
pytać	1	spóźniać	1	wbiec	1
realizować	1	spóźnić	1	wchodzić	1
relpron	1	ssać	1	wepchnąć	1
reprezentować	1	stosować	1	wetknąć	1
rodzić	1	stracić	1	widnieć	1
rozdeptać	1	strzec	1	wierzyć	1
rozgadywać	1	stukać	1	wieść	1
rozglądać	1	stwierdzić	1	włamać	1
rozmawiać	1	sypać	1	włożyć	1
rozmyślić	1	szanować	1	woleć	1
rozpoznać	1	szepnąć	1	wpadać	1
rozpuścić	1	tłumaczyć	1	wpaść	1
rozregulować	1	towarzyszyć	1	wpłynąć	1
rozstroić	1	tracić	1	wpływać	1
rozszerzać	1	traktować	1	wprowadzać	1
rozszerzyć	1	tupnąć	1	wręczać	1
roztaczać	1	tworzyć	1	wskazywać	1
rozumieć	1	uciekać	1	wstać	1
rysować	1	udawać	1	wstawać	1
rzucać	1	ukazywać	1	wtargnąć	1
sądzić	1	ukłuć	1	wybiec	1
scharakteryzować	1	umieć	1	wybierać	1
ściąć	1	umieścić	1	wybrać	1
ściągnąć	1	umrzeć	1	wybudować	1
ścisnąć	1	upić	1	wychylać	1
siedzieć	1	usiąść	1	wyciągnąć	1
sięgać	1	usłyszeć	1	wydobyć	1
składać	1	uspokajać	1	wydzierać	1
skończyć	1	ustać	1	wyjaśniać	1
skraść	1	uświadamiać	1	wykazać	1

wymachiwać	1	zacząć	1	zawadzać	1
wymagać	1	zaczynać	1	zawahać	1
wymierać	1	zadzwonić	1	zawężać	1
wymsknąć	1	zająknąć	1	zawierać	1
wynieść	1	zajść	1	zawołać	1
wypływać	1	zakasłać	1	zazgrzytać	1
wypowiedzieć	1	zakląć	1	zbić	1
wyprzeć	1	zakończyć	1	zbliżać	1
wyrażać	1	zaliczyć	1	zbliżyć	1
wyrecytować	1	zamarznąć	1	zdać	1
wyruszyć	1	zamienić	1	zdjąć	1
wyskoczyć	1	zamknąć	1	zdobyć	1
wystawiać	1	zamówić	1	zetknąć	1
wystrzelić	1	zamrozić	1	złocić	1
wysupłać	1	zamrzeć	1	zmęczyć	1
wysyłać	1	zanikać	1	zmienić	1
wytrzeć	1	zapatrzyć	1	znać	1
wytrzymać	1	zapewnić	1	znaczyć	1
wywołać	1	zapisać	1	znieść	1
wywrzeć	1	zapisywać	1	znikać	1
wzbogacać	1	zapłacić	1	zorganizować	1
wzbudzać	1	zaprojektować	1	zorientować	1
wzniecać	1	zaprosić	1	zostawiać	1
wznosić	1	zaprowadzić	1	zostawić	1
wzruszyć	1	zareagować	1	zraszać	1
zabierać	1	żartować	1	zrobić	1
zabijać	1	zaskoczyć	1	zwiększyć	1
zachęcić	1	zasłaniać	1	zwrócić	1
zachować	1	zasłuchać	1	zyskać	1
zachrypieć	1	zastanawiać	1		
zachwycać	1	zatrzepotać	1		
zaciągnąć	1	zauważyć	1		

Appendix G

Features used in the distance matrix

The features included in cluster analysis are:

- Temporal quantification:
  - SPEC – Specific temporal quantification
  - NONSPEC –Non-specific temporal quantification
  - PATTERN
  - STATEMENT
- PVA:
  - PFV – Perfective
  - IPFV – Imperfective
- FI_copula, PL_copula – Finnish/Polish copula
- simplex – Simplex Polish verb
- prefixed – Prefixed verb
- Polish prefixes
  - Z. – Prefix z-
  - ZA. – Prefix za-
  - PO. – Prefix po-
  - WY. – Prefix wy-

- Semantic macrolabel of subject
  - FI_Subj_static, PL_Subj_static – Finnish/Polish static subject (the combination of semantic labels zero, container, cogniser, positioned, content, stimulus, emoter, reference)
  - FI_Subj_agent, PL_Subj_agent – Finnish/Polish agent subject (the combination of semantic labels agent and speaker)
  - FI_Subj_positioner, PL_Subj_positioner – Finnish/Polish positioner subject (semantic label positioner)
  - FI_Subj_mover, PL_Subj_mover – Finnish/Polish positioner subject (semantic label mover)
  - FI_Subj_patient, PL_Subj_patient – Finnish/Polish patient subject (the combination of semantic labels medium, patient_qual, patient_quant)
  - FI_Subj_receiver, PL_Subj_receiver – Finnish/Polish receiver subject (the hybrid of semantic labels perceiver, recipient)
- Semantic macrolabel of object
  - FI_Obj_static, PL_Obj_static – Finnish/Polish object static (the combination of semantic labels container, reference, beneficiary, perceiver, emoter, topic, stimulus, content)
  - FI_Obj_mover, PL_Obj_mover – Finnish/Polish mover object (semantic label mover)
  - Finnish/Polish object patient (the combination of semantic labels medium, patient_qual, patient_quant)
  - FI_Obj_patient, PL_Obj_patient – Finnish/Polish patient subject (the combination of semantic labels medium, patient_qual, patient_quant)
- Finnish object types
  - FI_Obj_Tot – Total object
  - FI_Obj_Par – Partitive object
- Finnish oblique arguments
  - FI_Obl_Ade – Adessive argument
  - FI_Obl_All – Allative argument

  - FI_Obl_Ela – Elative argument
  - FI_Obl_Ess – Essive argument
  - FI_Obl_Ill – Illative argument
  - FI_Obl_Ine – Inessive argument
  - FI_Obl_Trans – Translative argument
- Polish oblique arguments
  - PL_Obl_Dat – Dative argument
  - PL_Obl_Ins – Instrumental argument
  - PL_Obl_prep_Acc – Prepositional Accusative argument
  - PL_Obl_prep_Gen – Prepositional Genitive argument
  - PL_Obl_prep_Loc – Prepositional Locative argument
- Temporal localising
  - PL_NPAST – Polish Non-past tense
  - PL_PAST – Polish Past tense
  - FI_NPAST – Finnish Present tense
  - FI_SPAST – Finnish Simple Past tense
  - FI_PERFECT – Finnish Perfect
  - FI_deictic_temp_localising, PL_deictic_temp_localising – Finnish/Polish lexical expression localising deictically
  - FI_absolute_temp_localising, PL_absolute_temp_localising – Finnish/Polish lexical expression localising to an absolute unit of time
  - FI_taxis_sequence, FI_taxis_sequence – Finnish/Polish clause localised temporally in sequence
  - FI_taxis_simultaneity, FI_taxis_sequence – Finnish/Polish clause localised temporally as simultaneous to another clause

Appendix H

Distance matrix

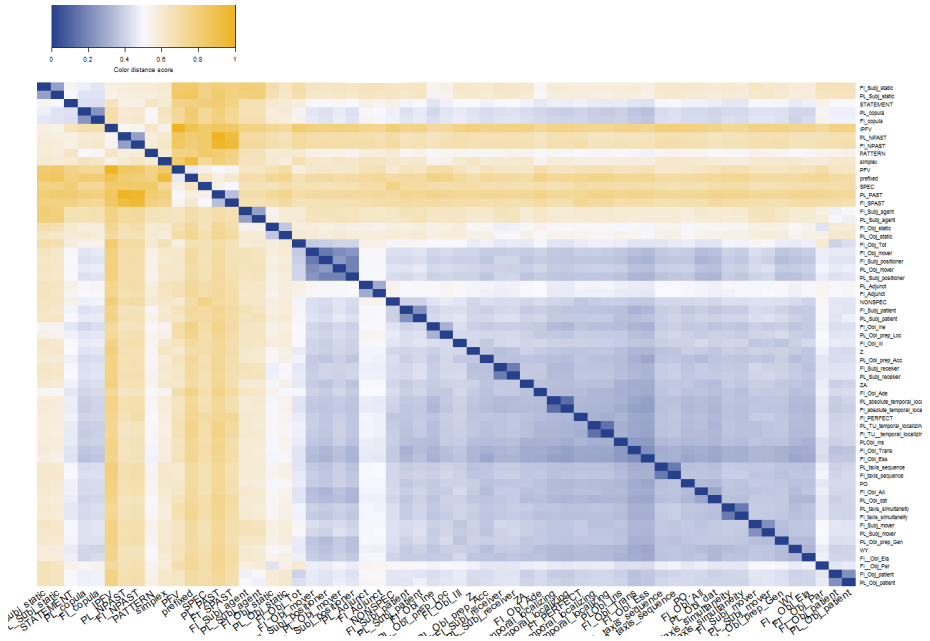

Figure H.1: Distance matrix represented as a heatmap

Appendix I

Distance coefficients for PFV and IPFV

PVA	SPEC	NON-SPEC	STATEMENT	PATTERN
PFV	**0.5520**	0.5963	0.7639	0.7763
IPFV	0.8338	0.8028	0.6454	0.6303

	PL SIMULTANEITY	PL SEQUENCE	FI SIMULTANEITY	FI SEQUENCE
PFV	0.6819	0.6084	0.6753	0.6066
IPFV	0.7314	0.7936	0.7376	0.7950

	PL NPAST	PL PAST	FI NPAST	FI PERF	FI SPAST
PFV	0.8506	**0.5150**	0.8466	0.6566	**0.5312**
IPFV	**0.5259**	0.8572	**0.5323**	0.7542	0.8473

	simplex	prefixed	PL copula	FI copula	PO-	Z-	ZA-	WY-
PFV	0.8345	**0.4606**	0.7291	0.7276	0.6195	0.6075	0.6213	0.6549
IPFV	**0.5510**	0.8876	0.6844	0.6860	0.7850	0.7943	0.7836	0.7557

	FI OBJ static	FI OBJ patient	FI OBJ mover	FI OBJ TOT	FI OBJ PAR
PFV	0.6660	0.6330	0.6140	**0.5691**	0.6918
IPFV	0.7459	0.7742	0.7893	0.8222	0.7221

	PL OBJ static	PL OBJ patient	PL OBJ mover				
PFV	0.6991	0.6312	0.6121				
IPFV	0.7150	0.7756	0.7907				
	FI SUBJ static	FI SUBJ mover	FI SUBJ agent	FI SUBJ patient	FI Subj receiver	FI SUBJ positioner	
PFV	0.8263	0.6330	0.6374	0.6488	0.6294	0.6213	
IPFV	**0.5632**	0.7742	0.7705	0.7609	0.7771	0.7836	
	PL SUBJ static	PL SUBJ mover	PL SUBJ agent	PL SUBJ patient	PL SUBJ receiver	PL SUBJ positioner	
PFV	0.8365	0.6383	0.6339	0.6312	0.6303	0.6222	
IPFV	**0.5479**	0.7698	0.7734	0.7756	0.7763	0.7829	
	FI OBL ELA	FI OBL ILL	FI OBL INE	FI OBL TRANS	FI OBL ADE	FI OBL ALL	FI OBL ESS
PFV	0.6514	0.6454	0.6975	0.6462	0.6827	0.6267	0.6575
IPFV	0.7587	0.7639	0.7166	0.7631	0.7307	0.7792	0.7535
	PL OBL INS	PL OBL dat	PL OBL PREP ACC	PL OBL PREP LOC	PL OBL PREP GEN		
PFV	0.6392	0.6471	0.6462	0.6934	0.6365		
IPFV	0.7690	0.7624	0.7631	0.7205	0.7712		

	PL deictic	PL absolute	FI deictic	FI absolute
PFV	0.6652	0.6392	0.6626	0.6410
IPFV	0.7467	0.7690	0.7490	0.7676

Bibliography

Agrell, Sigurd. 1908. *Aspektänderung und Aktionsartbildung beim polnischen Zeitworte: ein Beitrag zum Studium der indogermanische Präverbia und ihrer Bedeutungsfunktionen [the change of aspect and the notion of Aktionsart in Polish temporal words: a contribution to the study of Indogermanic preverbs and their functions of meanings]*. Lund: Håkan Ohlssons Buchdruckerei.

AllerMediaOy. 2014. Suomi 24 virkkeet -korpus (2016h2). `http://urn.fi/urn:nbn:fi:lb-2017021505`.

Antinucci, Francesco & Lucyna Gebert. 1957. Semantyka aspektu czasownikowego [the semantics of verbal aspect]. *Studia gramatyczne* 1. 7 – 43.

Askonen, Ebba. 2001. *Objektin aspektuaalinen sijanvalinta [aspectual choice of object case]*. PhD Dissertation: Oulun yliopisto.

Bacz, Barbara. 2005. For the unit of meaning of the Polish verbal prefix *za-*. In Adam Makkai, William J. Sullivan & Arle L. Rommel (eds.), *LACUS Forum XXXI: Interconnections*, 93 – 104. Huston, TX: University of Illinois at Chicago.

Baker, Mona. 1999. The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistics* 4(2). 281 – 298.

Bartnicka, Barbara, Björn Hansen, Wojtek Klemm, Volkmar Lehmann & Halina Satkiewicz. 2004. *Grammatik des Polnischen [the grammar of Polish]*. Munich: Verlag Otto Sagner.

Bartnicka, Barbara & Halina Satkiewicz. 2000. *Gramatyka języka polskiego: Podręcznik dla cudzoziemców [the grammar of Polish: Handbook for foreigners]*. Warsaw: Wiedza Powszechna.

Bertinetto, Pier Marco & Denis Delfitto. 2000. Aspect vs. actionality. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 189 – 226. Berlin: Mouton de Gruyter.

Biber, Douglas. 1989. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Liguistic Computing* 8(4). 243 – 257.

Biber, Douglas. 1995. *Dimensions of register variaton*. Cambridge: Cambridge University Press.

Biskupska, Anna Maija. 2018. *Verbi verbistä: puolan ja suomen johdetun verbileksikon merkitysrakenteen vertailua*. Helsinki: Unigrafia.

Błaszczak, Joanna, Patrycja Jabłońska, Dorota Klimek-Jankowska & Krzysztof Migdalski. 2014. The riddle of future tense in Polish: How much 'future' is there in 'future tense'? In Mikhail Kissine, Philippe De Brabanter & Saghie Sharifzadeh (eds.), *Future times / future tenses*, 165 – 204. Oxford: Oxford University Press.

BNC, Consortium. 2007. *The British national corpus, version 3 (bnc xml edition)*. Oxford: Bodleian Libraries University of Oxford. `http://www.natcorp.ox.ac.uk/`.

Bondarko, Alexander V. 1991. *Functional grammar: A field approach*. Amsterdam: John Benjamins Publishing Company.

Borer, Hagit. 2005. *The normal course of events*. Oxford, UK: Oxford University Press.

Borin, Lars, Markus Forsberg & Johan Roxendal. 2012. Korp – the corpus infrastructure of språkbanken. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of LREC 2012. istanbul: ELRA*, 474–478. Istanbul: European Language Resources Association.

Bossong, Georg. 1985. *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen [empirical study of universals: differential object marking in the New-Iranian languages]*. Tübingen: Narr.

Bralczyk, Jerzy (ed.). 2005. *Słownik 100 tysięcy potrzebnych słów [the dictionary of 100 thousand necessary words]*. Warszawa: Wydawnictwo Naukowe PWN.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45. 5 – 32.

Breiman, Leo, Joseph H. Friedman, Richard A. Olshen & Charles J. Stone. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

Caudal, Patrick. 2005. Degree scales and aspect. In Bart Hollebrandse, Angeliek Van Hout & Co Vet (eds.), *Cross linguistic views on tense, aspect and modality*, 103 –118. Amsterdam: Rodopi.

Chesterman, Andrew. 2004. Beyond the particular. In Anna Mauranen & Pekka Kujamäki (eds.), *Translation universals: do they exist?*, 33 – 49. Amsterdam: John Benjamins Publishing Company.

Chesterman, Andrew. 2010. Why study translational universals? In Ritva Hartama-Heinonen & Pirjo Kukkonen (eds.), *Acta translatologica helsingiensa*, 38 – 48. Helsinki: Kiasm.

Ciecieląg, Paweł, Mikołaj Haponiuk, Olga Lewandowska & Małgorzata Krzysztofik. 2013. *Wyznania religijne. stowarzyszenia narodowościowe i etniczne w Polsce 2009–2011 [churches, denominations as well as national and ethnic associations in Poland 2009–2011]*. Warszawa: Zakład Wydawnictw Statystycznych.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.

Croft, William. 2001. *Radical construction grammar*. Oxford: Oxford University Press.

Cysouw, Michael. 2007. Building semantic maps : The case of person marking. In Bernard Wälchli & Matti Miestamo (eds.), *New challenges in typology*, Berlin: Mouton de Gruyter.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.

Dahl, Östen (ed.). 2000. *Tense and aspect in the languages of Europe*. Berlin: Mouton de Gruyter.

Dahl, Östen. 2007. From questionnaires to parallel corpora in typology. *Sprachtypologie und Universalienforschung* 60(2). 172 – 181.

Dahl, Östen. 2012. Thoughts on language-specific and crosslinguistic entities. *Liguistic Typology* 20(2). 427 – 437.

Dahl, Östen & Bernhard Wälchli. 2016. Perfects and iamitives: Two gram types in one grammatical space. *Letras de Hoje* 51(3). 325 – 348.

Dickey, Stephen. 2000. *Parameters of Slavic aspect: A cognitive approach*. Stanford, Calif: CSLI Publications, Center for the Study of Language and Information.

Dik, Simon C. 1997. *The theory of functional grammar*. Berlin: Walter de Gruyter.

Dixon, Robert. 2010. *Basic linguistic theory: methodology*. Oxford: Oxford University Press.

Dąbrowska, Ewa. 1996. The spatial structuring of events: A study of Polish perfectivizing prefixes. In Martin Pütz & René Dirven (eds.), *The construal of space in language and thought*, 467 – 490. Berlin: Mouton de Gruyter.

Doroszewski, Witold (ed.). 1969. *Słownik języka polskiego [Polish dictionary]*. Warszawa: Polska Akademia Nauk.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 3(67). 547 – 619.

Dray, Stéphane & Anne-Béatrice Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22(4). 1 – 20.

Dryer, Matthew S. 1997. Are grammatical relations universal? In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type*, 115 –143. Amsterdam: Benjamins.

Ebeling, Jarle. 1998. Contrastive linguistics, translation and parallel corpora. *Meta* 43(4). 602 – 615.

Eskola, Sari. 2004. Untypical frequencies in translated language: A corpus based study on a literary corpus of translated and non-translated Finnish. In Anna Mauranen & Pekka Kujamäki (eds.), *Translation universals: do they exist?*, 83–99. Amsterdam: John Benjamins Publishing Company.

Fehrmann, Dorothee, Uwe Junghanns & Denisa Lenertová. 2010. Two reflexive markers in Slavic. *Russian Linguistics* 34. 203 – 238.

Filip, Hana. 1993. Verbal aspect and object case marking: A comparison between Czech and Finnish. In Joel Ashmore Nevis & Vida Samiian (eds.), *Proceedings of the Western Conference on Linguistics (WECOL) 22*, 43 – 59. Fresno: California State University.

Fillmore, Charles J. 1968. The case for case. In Emmon Bach & Robert Thomas Harms (eds.), *Universals in linguistic theory*, 1 – 88. New York: Holt, Rinehart, and Winston.

Garey, Howard B. 1957. Verbal aspect in French. *Language* 33(2). 91 – 110.

Gries, Stefan Th. 2013. *Statistics for linguistics with R: A practical introduction*. Berlin: De Gruyter Mouton.

Grönros, Eija-Riitta (ed.). 2018. *Kielitoimiston sanakirja [new dictionary of modern Finnish]*. Helsinki: Kotimaisten kielten keskus.

Groundstroem, Axel. 1988. *Finnische Kasusstudien [studies in Finnish case]*. Umeå: Universitetet i Umeå.

Grzegorczykowa, Renata, Roman Laskowski & Henryk Wróbel (eds.). 1984. *Współczesna gramatyka języka polskiego [the grammar of contemporary Polish]*. Warszawa: Wydawnictwo Naukowe PWN 1st edn.

Grzegorczykowa, Renata, Roman Laskowski & Henryk Wróbel (eds.). 1998. *Współczesna gramatyka języka polskiego [the grammar of contemporary Polish]*. Warszawa: Wydawnictwo Naukowe PWN 2nd edn.

Gvozdanović, Jadranka. 2012. Perfective and imperfective aspect. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 781 – 802. Oxford: Oxford University Press.

Hakulinen, Auli. 1987. Avoiding personal reference in Finnish. In Jef Verschueren & Marcella Bertuccelli Papi (eds.), *The pragmatic perspective: Selected papers from the 1985 international pragmatics conference*, 141 – 154. Amsterdam: John Benjamins Publishing Company.

Hakulinen, Auli & Fred Karlsson. 1979. *Nykysuomen lauseoppia [the syntax of modern Finnish]*. Jyväskylä: Suomalaisen Kirjallisuuden Seura.

Hakulinen, Auli, Fred Karlsson & Maria Vilkuna. 1980. *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus [the features of the Finnish sentences in text: a quantitative study]*. Helsinki: Helsingin yliopisto.

Hakulinen, Auli & Lauri Karttunen. 1973. Missing persons: On generic sentences in Finnish. In *Papers from the ninth regional meeting of the Chicago Linguistic Society*, 157 – 170. Chicago.

Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2004. *Iso suomen kielioppi [the great grammar of Finnish]*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Hardegree, Gary. 2001a. *Additive scales*. University of Massachusets Amherst. `https://people.umass.edu/gmhwww/382/pdf/11-additive%20scales.pdf`.

Hardegree, Gary. 2001b. *Measurement*. University of Massachusets Amherst. `https://people.umass.edu/gmhwww/382/pdf/10-measurement.pdf`.

Harrell, Frank & Charles Dupont. 2018. *Hmisc: harrell miscellaneous*. R package version 4.2-0.

Harris, Zellig. 1954. Distributional structure. *Word* 23(10). 146 – 162.

Haspelmath, Martin. 1997. *From space to time*. Munich: Lincom.

Haspelmath, Martin. 2009. Terminology of case. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 505 – 517. Oxford: Oxford University Press.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3). 663 – 687.

Haspelmath, Martin. 2015. Transitivity prominence. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, 131 –148. Berlin: de Gruyter Mouton.

Haspelmath, Martin. 2016. The interplay between comparative concepts and descriptive categories (reply to Newmeyer). *Language* 86(3). 696 – 699.

Haspelmath, Martin & Andrea D. Sims. 2010. *Understanding morphology*. London: Hodden Education 2nd edn.

Haverinen, Katri, Jenna Nyblom, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Anna Missilä, Stina Ojala, Tapio Salakoski & Filip Ginter. 2014. Building the essential resources for Finnish: the Turku dependency treebank. *Language Resources and Evaluation* 48. 493 – 531.

Heinämäki, Orvokki. 1984. Aspect in Finnish. In Casper de Groot & Hannu Tommola (eds.), *Aspect bound: a voyage into the realm of Germanic, Slavonic and Finno-Ugrian aspectology*, 153–178. Dordrecht: Foris Publications.

Heinämäki, Orvokki. 1994. Aspect as boundedness in Finnish. In Carl Bache, Hans Basbøll & Carl-Erik Lindberg (eds.), *Tense, aspect and action: empirical and theoretical contributions to language typology*, 207 – 234. Berlin: De Gruyter.

Hertzen-Oosi von, Nina, Henna Herju, Niina Haake & Aro Timo. 2009. *Ulkomaalaisten tilapäinen työnteko Suomessa [the temporary labour of foreigners in Finland]*. Helsinki: Työ- ja elinkeinoministeriö.

Holvoet, Axel. 1989. *Aspekt a modalność w języku polskim: na tle ogólnosłowiańskim [aspect and modality in Polish: in the pan-Slavic context]*. Wrocław: Zakład Narodowy im. Ossolińskich.

Hopper, Paul & Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251 – 299.

Hothorn, Torsten, Peter Buehlmann, Sandrine Dudoit, Annette Molinaro & Mark J. Van Der Laan. 2006a. Survival ensembles. *Biostatistics* 7. 355 – 373.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006b. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15. 651 – 674.

Huumo, Tuomas. 1997. *Lokatiivit lauseen semanttisessa tulkinnassa: ajan, omistajan, paikan ja tilan adverbiaalien keskinäiset suhteet suomen kielessä [locatives and semantic interpretation of the sentence: on mutual relations of adverbials indicating time, possession, space and (internal) state in Finnish]*. Turku: Turun yliopisto.

Huumo, Tuomas. 2006. Kvantiteetti ja aika I: Nominaalisen aspektin näkökulma suomen objektin ja subjektin sijanmerkintään [quantity and time I: nominal aspect and the case marking of subject and object in Finnish]. *Virittäjä* 110(4). 504–538.

Huumo, Tuomas, Aki-Juhani Kyröläinen, Jenna Kanerva, M. Juhani Luotolahti, Tapio Salakoski, Filip Ginter & Veronika Lappala. 2017. Distributional semantics of partitive A-argument constructions. In Milla Luodonpää-Manni, Esa Esa Penttilä & Johanna Viimaranta (eds.), *Empirical approaches to cognitive linguistics: Analysing real-life data*, 25 – 48. Cambridge: Cambridge Scholars Publishing.

Huumo, Tuomas & Krista Ojutkangas. 2006. An introduction to Finnish spatial relations. In Marja-Liisa Helasvuo & Lyle Campbell (eds.), *Grammar from the human perspective: Case, space and person in Finnish*, 11 – 20. Amsterdam: John Benjamins Publishing Company.

Hynönen, Emmi. 2017. The essive in Finnish. In Casper de Groot (ed.), *Uralic essive and the expression of impermanent state*, 29 – 56. Amsterdam: John Benjamins Publishing Company.

Ilves, J.M. 2016. *Sorjonen: nukkekoti*. Juva: Gummerus.

Itkonen, Terho. 1976. *Erään sijamuodon ongelmia [the problems of one case]*. Helsinki: Helsingin yliopisto.

Jakobson, Roman. 1957[1971]. Shifters, verbal categories and the Russian verb. In *Selected writings*, vol. II, 130 – 147. The Hague: Mouton de Gruyter.

Jakobson, Roman. 1971[1932]. Zur Struktur des russichen Verbums. In *Selected writings*, vol. II, 3 – 15. The Hague: Mouton de Gruyter.

Johansson, Stig. 2007. Seeing through multilingual corpora. In Roberta Facchinetti (ed.), *Corpus linguistics 25 years on*, 51 – 72. Amsterdam: Rodopi.

Jurkiewicz-Rohrbacher, Edyta, Zrinka Kolaković & Björn Hansen. 2017. Web Corpora – the best possible solution for tracking rare phenomena in underresourced languages: clitics in Bosnian, Croatian and Serbian. In Piotr Bański, Marc Kupietz, Harald Lüngen, Paul Rayson, Hanno Biber, Evelyn Breiteneder, Simon Clematide, John Mariani, Mark Stevenson & Theresa Sick (eds.), *Proceedings of the workshop on challenges in the management of large corpora and big data and natural language processing (cmlc-5+bignlp) 2017 including the papers from the web-as-corpus (wac-xi) guest section*, 49 – 55. Mannheim: Institut für Deutsche Sprache.

Káňa, Tomáš & Hana Pelouškova. 2010. *Česko-německý paralelní korpus*. Masarykova univerzita. `https://www.ped.muni.cz/katedry-a-instituty/nemecky-jazyk-literatura/aktivity/cesko-nemecky-paralelni-korpus`.

Kangasmaa-Minn, Eeva. 1978. Verbien sisäisestä aspektista [On the hidden aspect of Finnish verbs]. *Sananjalka* 20. 17 – 28.

Kangasmaa-Minn, Eeva. 1984. Tense, aspect and Aktionsarten in Finno-Ugrian. In Casper De Groot & Hannu Tommola (eds.), *Aspect bound: A voyage into the realm of Germanic, Slavonic and Finno-Ugrian aspectology*, 77 – 96. Dordrecht: Foris Publications.

Kangasmaa-Minn, Eeva. 1999. *Mitä tulikaan sanotuksi: Omakohtaisia kielitieteellisiä oivalluksia [what was said: personal linguistic thoughts]*. Turku: Publications of the Department of Finnish and General Linguistics of the University of Turku.

Kenny, Dorothy. 1998. Creatures of habit? What translators usually do with words. *Meta* 43(4). 515 – 523.

Kiefer, Ferenc & Johanna Laakso. 2014. Uralic. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*, 473 – 492. Oxford: Oxford University Press.

Kilgariff, Adam, Michael Rundell & Elaine Uí Dhonnchadha. 2006. Efficient corpus development for lexicography: building the new corpus for Ireland. *Language resources and evaluation* 40(2). 127 – 152.

Kiparsky, Paul. 2005. Absolutely a matter of degree: The semantics of structural case in Finnish. Handout for CLS.

Kochańska, Agata. 2002. A cognitive grammar analysis of Polish nonpast perfectives and imperfectives: How virtual events differ from actual ones. In Frank Brisard (ed.), *Grounding*, 349 – 390. Berlin: De Gruyter Mouton.

Kochańska, Agata. 2007. Conflicting epistemic meanings of the Polish aspectual variants. In Agata Kochańska & Dagmar Divjak (eds.), *Cognitive paths into the Slavic domain*, 149 – 180. Berlin: De Gruyter Mouton.

Koivisto, Vesa. 2013. *Suomen sanojen rakenne [the structure of words in Finnish]*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kolehmainen, Leena. 2005. *Präfix- und Partikelverben im deutsch-finnischen Kontrast [prefix and particle verbs in the German-Finnish contrast]*. Frankfurt am Main: Peter Lang.

Koschmieder, Erwin. 1934. *Nauka o aspektach czasownika polskiego w zarysie: Próba syntezy [the study of Polish verbal aspects: synthesis]*. Wilno: Towarzystwo Przyjaciół Nauk w Wilnie.

Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In Ivan Sag & Anna Szabolcsi (eds.), *Lexical matters*, 29 – 53. Stanford: CSLI Publications.

Krifka, Manfred, Francis J. Pelletier, Gregory Carlson, Alice ter Meulen, Gennaro Chierchia & Godehard Link. 1995. Genericity: An introduction. In Gregory Carlson & Francis J. Pelletier (eds.), *The generic book*, 1 – 124. Chicago: University of Chicago Press.

Kroeger, Paul R. 2004. *Analyzing syntax: A lexical-functional approach*. Cambridge: Cambridge University Press.

Krzywacki, Jyrki & Kaisa Saarenmaa. 2013. Suomessa työskentelee jo lähes 60 000 vierastyöläistä. *Hyvinvointikatsaus* 3(23). `https://www.stat.fi/artikkelit/2013/art_2013-09-23_013.html`.

Laakso, Johanna. 1997. On verbalising nouns in Uralic. *Finnisch-Ugrische Forschungen* 54(3). 267 – 304.

Larjavaara, Matti. 2007. *Pragmasemantiikka [pragmasemantics]*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Laskowski, Roman. 1996. Aspekt a znaczenie czasowników (predykaty zmiany stanu) [aspect and verbal semantics (predicates of the change of state)]. In Ewa Bednarska-Gryniewicz (ed.), *Studia z leksykologii i gramatyki języków słowiańskich. IV polsko-szwedzka konferencja slawistyczna Mogilany, 1-3 października 1995*, 43 – 59. Kraków: Wydawnictwo Instytutu Języka Polskiego PAN.

Laskowski, Roman. 1998a. Aspekt czasowników przemieszczania (ukierunkowanego ruchu liniowego)[the aspect of the verbs of motion (directed linear motion)]. In Janusz Siatkowski (ed.), *Prace na XII międzynarodowy kongres slawistów w Krakowie 1998*, 43 – 59. Warszawa: Wydawnictwo Energeia.

Laskowski, Roman. 1998b. Czasownik [verb]. In Renata Grzegorczykowa, Roman Laskowski & Henryk Wróbel (eds.), *Współczesna gramatyka języka polskiego [the grammar of contemporary Polish*, vol. I, 225 – 269. Warszawa: Wydawnictwo Naukowe PWN 2nd edn.

Laskowski, Roman. 2003. Präpositionale Ausdrücke mit temporaler Funktion im Polnischen [polish prepositional expressions with temporal function]. In Gerd Hentschel & Thomas Mentzel (eds.), *Präpositione im Polnischen*, 193 – 226. Oldenburg: Biblioteks- und Informationssystem der Universität Oldenburg.

Łaziński, Marek. 2011. Polish aspectual prefixes, their order and functions: A study based on the National Corpus of Polish. *Word Structure* 4(2). 231–243.

Łazorczyk, Agnieszka. 2010. *Decomposing Slavic aspect: the role of aspectual morphology in Polish and other Slavic languages*. PhD Dissertation: University of Southern California.

Lehmann, Volkmar. 2009. Aspekt und Tempus in Slavischer. In Tilman Berger, Karl Gutschmidt, Sebastian Kempgen & Peter Kosta (eds.), *Slavische Sprachen – Slavic languages*, 526 – 556. Berlin: Walter de Gruyter.

Leino, Jaakko. 2010. Results, cases and constructions: Argument structure constructions in English and Finnish. In Hans C. Boas (ed.), *Contrastive studies in construction grammar*, 103 – 135. Berlin: John Benjamins Publishing Company.

Leino, Pentti. 1991. *Lauseet ja tilanteet: suomen objektin ongelmia [sentences and situations: the problems related to the Finnish object]*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Publishing Company.

Lin, Jo-Wang. 2012. Tenselessness. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 669 – 695. Oxford: Oxford University Press.

Lindstedt, Jouko. 1984. Nested aspects. In Casper De Groot & Hannu Tommola (eds.), *Aspect bound: A voyage into the realm of Germanic, Slavonic and Finno-Ugrian aspectology*, 23 – 38. Dordrecht: Foris Publications.

Lindstedt, Jouko. 1995. Understanding perfectivity – understanding bounds. In Pier Marco Bertinetto, Valentina Bianchi, Östen Dahl & Mario Squartini (eds.), *Temporal reference, aspect and actionality : Typological perspectives*, vol. 2, 95–103. Torino: Rosenberg and Sellier.

Lindstedt, Jouko. 2001. Tense and aspect. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfagang Raible (eds.), *Language typology and language universals*, vol. 1, 768–783. Berlin: Walter de Gruyter.

Lingorska, Blagovesta. 1977. Funkcionalno-semantični săotvetstvija na bălgarskija pluskvamperfekt v polski ezik [the functional-semantic correlations of the Bulgarian pluperfect in Polish]. In Ivan Lekov (ed.), *Zakonomernosti na razvitieto na slavjanskite ezici*, 149 – 167. Sofiia: BAN.

Łuczków, Iwona. 2013. O semantyce taxis (na materiale języka polskiego i rosyjskiego) [about semantics of taxis on the Polish and Russian material]. *Studia Linguistica Polono-(Meridiano)Slavica* 14–15. 91 – 105.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert & Kurt Hornik. 2018. *cluster: Cluster analysis basics and extensions*. R package version 2.0.7-1.

Maslov, Jurij S. 1978. K osnovanijam sopostavitel'noj aspektologii [on the foundations of contrastive aspectology]. In Jurij S. Maslov (ed.), *Voprosy sopostavitel'noj aspektologii*, 4 – 44. Leningrad: Izdatel'stvo leningradskogo universiteta.

McEnery, Tony, Richard Xiao & Yukio Tono. 2006. *Corpus-based language studies*. London: Routledge.

Meyer, David, Achim Zeileis & Kurt Hornik. 2017. *vcd: Visualizing categorical data*. R package version 1.4-4.

Mihalcea, Rada & Michel Simard. 2005. Parallel texts. *Journal of Natural Language Engineering* 11(3). 239 – 246.

Mourelatos, Alexander P.D. 1978. Events, processes and states. *Linguistics and Philosphy* 2. 415 – 434.

Newman, Paul. 1980. *The classification of Chadic within Afroasiatic*. Leiden: Universitare Press.

Newmeyer, Frederick J. 2007. Linguistic typology requires crosslinguistic formal categories. *Linguistic Typology* 110. 133 – 157.

Nowak, Lucyna (ed.). 2012. *Raport z wyników: Narodowy spis powszechny ludności i mieszkań 2011 [results report: Polish census of 2011]*. Warszawa: Zakład Wydawnictw Statystycznych.

Næss, Åshild. 2008. Varieties of dative. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 339 – 355. Oxford: Oxford university press.

Nurminen, Salla. 2011. *Lauseen rajatun ja rajaamattoman aspektin määräytyminen : kielen käytön näkökulma*. Turku: Turun yliopisto.

Nurminen, Salla. 2014. Aspektin ilmaiseminen suomen kielessä – lauseen ominaisuus [expressing aspect in Finnish: a property of the clause]. *Sananjalka* 53. 61 – 83.

Nurminen, Salla. 2015. *Lauseen aspektin ilmaiseminen suomen kielessä : tutkimus olla-, tehdä- ja tulla-verbillisistä tekstilauseista käyttökonteksteissaan [expressing clausal aspect in Finnish: A study of clauses with the verbs olla, tehdä and tulla in their textual contexts]*. PhD Dissertation: Turun yliopisto.

Nurminen, Salla. 2017. A usage-based and contextual approach to clausal aspect in Finnish. In Milla Luodonpää-Manni, Esa Esa Penttilä & Johanna Viimaranta (eds.), *Empirical approaches to cognitive linguistics: Analysing real-life data*, 150 – 175. Cambridge: Cambridge Scholars Publishing.

Päiviö, Pia. 2007. *Suomen kielen asti ja saakka: Terminatiivisten partikkelien synonymia, merkitys, käyttö ja kehitys sekä asema kieliopissa [the Finnish asti and saakka: the synonymy, meaning, use, development and grammatical category of terminative particles]*. PhD Thesis: Turun yliopisto.

Perlin, Jacek. 2010. Ile jest we współczesnej polszczyźnie czasowników dwuaspektowych? [how many bi-aspectual verbs are there in modern Polish?]. *Linguistica Copernicana* 3. 165 – 171.

Pleciński, Marek, Maciej Czerniakiewicz & Paweł Błaszczyk. 2014. *Finlandia – bliżej niż myślisz: przewodnik dla polskich handlowców i inwestorów*. Helsinki: Wydział promocji handlu i inwestycji w Helsinkach.

Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górski & Barbara Lewandowska-Tomaszczyk (eds.). 2012. *Narodowy korpus języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.

Radziszewski, Adam. 2013. A tiered CRF tagger for Polish. In Ryszard Bembenik, Łukasz Skonieczny, Henryk Rybiński, Marzena Kryszkiewicz & Marek Niezgódka (eds.), *Intelligent tools for building a scientific information platform: Advanced architectures and solutions*, 215 – 230. Springer Verlag.

Rand, William. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* 66(336). 846 – 850.

Reppen, Randi. 2010. Building a corpus: What are the key considerations? In O'keeffe Anne & Michael McCarthy (eds.), *The Routledge handbook of corpus linguisitics*, 31 – 37. New York: Routledge.

Rousseeuw, Peter. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20. 53 – 65.

Sahlgren, Magnus. 2008. The distributional hypothesis. *Rivista di Linguistica (Italian Journal of Linguistics)* 1(20). 33 – 53.

Santos, Diana. 1996. *Tense and aspect in English and Portuguese: a contrastive semantical study*. PhD Dissertation: Universidade Tecnica de Lisboa.

Shluinsky, Andrey. 2009. Individual-level meanings in the semantic domain of pluractionality. In Patience Epps & Alexandre Arkhipov (eds.), *New challenges in typology: Transcending the borders and refining the distinctions*, 175 – 197. Berlin: Mouton de Gruyter.

Siro, Paava. 1964. *Suomen kielen lauseoppi*. Helsinki: Tietosanakirja Oy.

Sivonen, Jari. 2007. Aspektin ontologia ja kognitiivinen semantiikka [the ontology of aspect and cognitive semantics]. In Harri Mantila (ed.), *Merkityksen ongelmasta vähemmistökielten oikeuksiin: juhlakija professorin Helena Sulkalan 60-vuotispäivänä*, 73 – 95. Oulun yliopisto.

Śmiech, Witold. 1971. *Funkcje aspektów czasownikowych we współczesnym języku polskim [the functions of the verbal aspects in the contemporary Polish language]*. Łódź: Łódzkie Towarzystwo Naukowe.

Śmiech, Witold. 1986. *Derywacja prefiksalna czasowników polskich [the prefixal derivation of Polish verbs]*. Wrocław: Łódzkie Towarzystwo Naukowe.

Smith, Carlota S. 2003. *Modes of discourse : the local structure of text*. Cambridge: Cambridge University Press.

Sokal, Robert R. & Charles D. Michener. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38(22). 1409 – 1438.

Stawnicka, Jadwiga. 2007. *Aspekt – iteratywność – określniki kwantyfikacyjne (na materiale form czasu przeszłego w języku polskim) [aspect – iteration – quantifying determiners (on the basis of the past tense in Polish)]*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Stevens, Stanley Smith. 1946. On the theory of scales of measurement. *Science* 103(2684). 677 – 680.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9. 307.

Stunová, Anna. 1993. *A contrastive study of Russian and Czech aspect: invarianve vs. discourse*. PhD Dissertation: Universiteit van Amsterdam.

Suzuki, Ryota & Hidetoshi Shimodaira. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12). 1540 – 1542.

Szal, Katarzyna. 2013. *Finnish literature in Poland, Polish literature in Finland: comparative reception study from a hermeneutic perspective*. PhD Dissertation: University of Eastern Finland.

Tabakowska, Elżbieta. 2003. Space and time in Polish: the verbal preposition za and the verbal prefix za-. In Hubert Cuyckens et (ed.), *Motivation in language: Language studies in honor of Günter Radden*, 53 — 77. Amsterdam: John Benjamins.

Tagliamonte, Sali A. & Harald Baayen. 2012. Models forests and trees of York English: Was/were variations a case study for statistical practice. *Language variation and change* 24(2). 135 – 178.

Tamm, Anne. 2012. *Scalar verb classes: Scalarity, thematic roles, and arguments in the Estonian aspectual lexicon*. Florence: Firenze University Press.

Tenny, Carol. 1994. *Aspectual roles and the syntax-semantics interface*. Dordrecht: Kluwer Academic Publishers.

Tiedemann, Jörg. 2012. Parallel data, tools and interfaces in OPUS. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *In proceedings of the 8th international conference on language resources and evaluation (LREC'2012)*, 2214 – 2218. Istanbul: European Language Resources Association (ELRA).

Tilastokeskus (ed.). 2014. *Väestörakenne 2013 [the structure of population 2013]*. Helsinki: Edita Publishing.

Tirkkonen-Condit, Sonja. 2004. Unique items – over- or under-represented in translated language? In Anna Mauranen & Pekka Kujamäki (eds.), *Translation universals : do they exist?*, 177 – 184. Amsterdam: John Benjamins Publishing Company.

Tommola, Hannu. 1986. *Aspektual'nost' v finskom i russkom jazykah [aspectuality in Finnish and Russian]*. Helsinki: Neuvostoliittoinstituutti.

Tommola, Hannu. 1990. On Finnish "aspect" in discourse. In Nils B. Thelin (ed.), *Verbal aspect in discourse*, 349 – 364. Amsterdam: John Benjamins Publishing Company.

Tommola, Hannu. 2016. Taxis in Finnish. In Viktor S. Chrakovskij (ed.), *Typology of taxis constructions*, 385 – 427. Munich: Lincom.

Toury, Gideon. 1995. *Descriptive translation studies and beyond*. Amsterdam: John Benjamins.

Tuomikoski, Risto. 1978. Objektinsijaisista määrän adverbiaaleista [measure adverbials in the object cases]. *Virittäjä* 1. 26 – 51.

Van Valin, Robert D. 1999. Generalized semantic roles and the syntax-semantics interface. In Francis Corblin, Carmer Dobrovie-Sorin & Jean-Marie Marandin (eds.), *Empirical issues in formal syntax and semantics*, 373–389. The Hague: Thesus.

Čermák, František & Aleksandr Rosen. 2012. The case of Intercorp, a multilingual parallel corpus. *International Journal of Corpus Liguistics* 3(17). 411 – 427.

Vendler, Zeno. 1957. Verbs and times. *Philosophical review* 46. 143 – 160.

Vilkuna, Maria. 1992. *Referenssi ja määräisyys suomenkielisten tekstien tulkinnassa [reference and definiteness in the interpretation of Finnish texts]*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Vilkuna, Maria. 2000. *Suomen lauseopin perusteet [the basics of Finnish syntax]*. Helsinki: Edita.

Wälchli, Bernhard. 2007. Advantages and disadvantages of using parallel texts in typological investigations. *Sprachtypologie und Universalienforschung* 60(2). 95 – 199.

Ward, Joe H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58. 236 – 244.

Werlich, Egon. 1976. *A text grammar of English*. Heidelberg: VERLAG.

Whaley, Marika Lynn. 2000. *The evolution of the Slavic 'BE(COME)'-type compound future*. PhD Dissertation: The Ohio State University.

Wiemer, Björn & Ilja A. Seržant. 2017. Diachrony and typology of Slavic aspect: What does morphology tell us? In Walter Bisang & Andrej Malchukov (eds.), *Unity and diversity in grammaticalization scenarios*, 239 – 307. Berlin: Language Science Press.

Wierzbicka, Anna. 1967. On the semantics of verbal aspect in Polish. In *To honour Roman Jakobson*, 2231 – 2249. The Hague: Mouton.

Wróbel, Henryk. 1998. Czasownik [verb]. In Renata Grzegorczykowa, Roman Laskowski & Henryk Wróbel (eds.), *Współczesna gramatyka języka polskiego [the grammar of contemporary Polish*, vol. II, 536 – 584. Warszawa: Wydawnictwo Naukowe PWN 2nd edn.

Xiao, Richard & Anthony Mark McEnery. 2006. A corpus-based approach to tense and aspect in English-Chinese translation. In Johannes Schwittalla & Werner Wegstein (eds.), *Korpuslinguistik deutsch: synchron - diachron - konstrastiv*, 27 – 50. Tübingen: Niemeyer.

Żmigrodzki, Piotr (ed.). 2007. *Wielki słownik języka polskiego PAN [the great Polish dictionary of the Polish Academy of Science]*. Kraków: Instytut Języka Polskiego PAN.

Zmrzlíková, Jana. 2009. *Vyjadřování aspektu v češtině a ve finštině [the expression of aspect in Czech and Finnish]*. M.A. thesis: Masarykova univerzita, Filozofická fakulta.